```
OM of: US-09-988-687-2 to: Issued_Patents_NA:*   out_format : pfs

Date: Mar 30, 2002 12:54 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09988687/runat_29032002_092149_591/app_query.fasta_1.895
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=olip2n.rni
 -GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
 -XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
 -YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09988687_@CGN1_1_81 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-988-687-2
Query length: 826
Database: Issued_Patents_NA:*
Database sequences: 351203
Database length: 113238999
Search time (sec): 53.500000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence                Strd Orig    ZScore    EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-328-111-315 +    78.00 1410.43  6.5e-71   238  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-315-794-51 +     9.00  127.00  19.99     2517  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-389-341-51 +     9.00  127.00  19.99     2517  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-578-551-18 +     8.00  124.82  26.45     295  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-190-982-18 +     8.00  124.82  26.45     295  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-408-257-18 +     8.00  124.82  26.45     295  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-329-350-30 -     8.00  116.11  80.76     936  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-482-918-60 +     8.00  113.06  119.52    1404  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-224-681-60 +     8.00  113.06  119.52    1404  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-336-728A-60 +    8.00  113.06  119.52    1404  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-035-706-1 +      8.00  111.23  151.07    1789  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-955-841-1 +      8.00  111.23  151.07    1789  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-669-536-1 +      8.00  111.13  152.95    1812  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-929-580B-1 -     8.00  111.03  154.91    1836  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-929-580B-4 -     8.00  111.02  155.15    1839  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-379-496-1 -      8.00  109.78  182.00    2169  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-317-522A-1 +     8.00  108.64  210.72    2524  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-439-818A-1 +     8.00  108.64  210.72    2524  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-751-965-1 +      8.00  108.64  210.72    2524  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-738-975-1 +      8.00  108.64  210.72    2524  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-728-626-1 +      8.00  108.64  210.72    2524  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-808-599A-1 +     8.00  108.64  210.72    2524  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-591-025-8 -      8.00  108.44  216.21    2592  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-482-918-47 +     8.00  102.28  476.07    5864  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-224-681-47 +     8.00  102.28  476.07    5864  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-336-728A-47 +    8.00  102.28  476.07    5864  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-813-819-3 -      8.00   94.20  1.3e+03   17138  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2 -     8.00   52.37  2.3e+05   4403765
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-434-001-50 +     7.00  117.00  72.06     73  !  S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-433-585-50 +     7.00  117.00  72.06     73  !  S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-434-425-50 +     7.00  117.00  72.06     73  !  S
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-437-667-50 +     7.00  117.00  72.06     73  !  S
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-906-955-50 +     7.00  117.00  72.06     73  !  S
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-945-909-50 +     7.00  117.00  72.06     73  !  S
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US96-06060-50 +    7.00  117.00  72.06     73
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-606-789-9 -      7.00  112.31  131.50    136  !  S
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-111-348-9 -      7.00  112.31  131.50    136  !  S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-606-789-25 -     7.00  110.24  171.51    179  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-111-348-25 -     7.00  110.24  171.51    179
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-606-789-18 -     7.00  110.03  176.14    184
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-111-348-18 -     7.00  110.03  176.14    184
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-330-108-16 +     7.00  109.48  189.08    198
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US92-10087-16 +    7.00  109.48  189.08    1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-606-789-19 -     7.00  108.25  221.30    233
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-111-348-19 -     7.00  108.25  221.30    233

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-328-111-315

seq_documentation_block:
; Sequence 315, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
;  APPLICANT: Endege, Wilson O.
;  APPLICANT: Steinmann, Kathleen E.
;  APPLICANT: Astle, Jon H.
;  APPLICANT: Burgess, Christopher C.
;  APPLICANT: Bushnell, Steven E.
;  APPLICANT: Carroll III, Eddie
;  APPLICANT: Catino, Theodore J.
;  APPLICANT: Derti, Adnan
;  APPLICANT: Ford, Donna M.
;  APPLICANT: Lewis, Marcia E.
;  APPLICANT: Monahan, John E.
;  APPLICANT: Schlegel, Robert
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS
;  FILE REFERENCE: CCD-257 (US)
;  CURRENT APPLICATION NUMBER: US/09/328,111
;  CURRENT FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: US 60/088,801
;  EARLIER FILING DATE: 1998-06-10
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 315
;   LENGTH: 238
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-328-111-315


alignment_scores:
            Quality:   78.00              Length:     78
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-328-111-315   ..

 Align seg 1/1  to: US-09-328-111-315  from: 1  to: 238

      60 LeuGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrVa 76
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CTGCAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGT 52

      76 lPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGlyValGlnA 93
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 CTTCTCCGAGTTCAACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGA 102

      93 rgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePhe 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 GACTCATGCAGGAGCACAAGTTAAAGGTTGCTCGCCTGGACAACATATTC 152

     110 LeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIleLe 126
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 CTGACACGAATGCACTGGTCTAATGTTGGGGGCTTAAGTGGAATGATTCT 202

     126 uThrLeuLysGluThrGlyLeuProLysCysVal 137
         |||||||||||||||||||||||||||||||||
     203 TACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTA 236

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-315-794-51
```

```
seq_documentation_block:
; Sequence 51, Application US/09315794
; Patent No. 6197517
; GENERAL INFORMATION:
;  APPLICANT: Roberts, Christopher J.
;  TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
;  TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
;  TITLE OF INVENTION: DRUGS
;  FILE REFERENCE: 9301-053
;  CURRENT APPLICATION NUMBER: US/09/315,794
;  CURRENT FILING DATE: 1999-05-21
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 2517
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-315-794-51


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-315-794-51   ..

 Align seg 1/1  to: US-09-315-794-51  from: 1  to: 2517

     682 LeuLeuIleHisGluAlaThrLeuGlu 690
         ||||||||||||||||||||||||||||
    2149 CTATTAATTCACGAAGCTACACTAGAA 2175

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-389-341-51

seq_documentation_block:
; Sequence 51, Application US/09389341
; Patent No. 6200803
; GENERAL INFORMATION:
;  APPLICANT: Roberts, Christopher J.
;  TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
;  TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATIVE
;  TITLE OF INVENTION: DRUGS
;  FILE REFERENCE: 9301-057
;  CURRENT APPLICATION NUMBER: US/09/389,341
;  CURRENT FILING DATE: 1999-09-02
;  EARLIER APPLICATION NUMBER: 09/315,794
;  EARLIER FILING DATE: 1999-05-21
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 2517
;   TYPE: DNA
;   ORGANISM: Saccharomyces cerevisiae
US-09-389-341-51


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-389-341-51   ..

 Align seg 1/1  to: US-09-389-341-51  from: 1  to: 2517

     682 LeuLeuIleHisGluAlaThrLeuGlu 690
         ||||||||||||||||||||||||||||
    2149 CTATTAATTCACGAAGCTACACTAGAA 2175
```

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-578-551-18

seq_documentation_block:
; Sequence 18, Application US/08578551
; Patent No. 5854050
;  GENERAL INFORMATION:
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Sakari M.
;    APPLICANT:  Nielsen, Jack B.
;    APPLICANT:  Dambmann, Claus
;    TITLE OF INVENTION:  An Enzyme with Protease Activity
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5854050o No. 5854050disk of No. 5854050th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/578,551
;      FILING DATE:  01-FEB-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DK 0811/93
;      FILING DATE:  06-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 95/02044
;      FILING DATE:  19-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4006.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus aculeatus
US-08-578-551-18


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-08-578-551-18   ..

 Align seg 1/1  to: US-08-578-551-18  from: 1  to: 295

     265 ProValGlyThrAlaAlaIleAla 272
         ||||||||||||||||||||||||
     114 CCTGTTGGCACCGCTGCTATTGCT 137

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-190-982-18
```

```
seq_documentation_block:
; Sequence 18, Application US/09190982
; Patent No. 5998190
;  GENERAL INFORMATION:
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Sakari M.
;    APPLICANT:  Nielsen, Jack B.
;    APPLICANT:  Dambmann, Claus
;    TITLE OF INVENTION:  An Enzyme with Protease Activity
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 5998190o No. 5998190disk of No. 5998190th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/190,982
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/578,551
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 95/02044
;      FILING DATE:  19-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4006.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus aculeatus
US-09-190-982-18


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-190-982-18   ..

 Align seg 1/1  to: US-09-190-982-18  from: 1  to: 295

     265 ProValGlyThrAlaAlaIleAla 272
         ||||||||||||||||||||||||
     114 CCTGTTGGCACCGCTGCTATTGCT 137

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-408-257-18
```

```
seq_documentation_block:
; Sequence 18, Application US/09408257
; Patent No. 6190905
;  GENERAL INFORMATION:
;    APPLICANT:  Dalboge, Henrik
;    APPLICANT:  Christgau, Stephan
;    APPLICANT:  Andersen, Lene N.
;    APPLICANT:  Kofod, Lene V.
;    APPLICANT:  Kauppinen, Sakari M.
;    APPLICANT:  Nielsen, Jack B.
;    APPLICANT:  Dambmann, Claus
;    TITLE OF INVENTION:  An Enzyme with Protease Activity
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6190905o No. 6190905disk of No. 6190905th America, Inc.
;      STREET:  405 Lexington Avenue, 64th Floor
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10174-6401
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/408,257
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/578,551
;      FILING DATE:  01-FEB-1996
;      APPLICATION NUMBER:  DK 0811/93
;      FILING DATE:  06-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 95/02044
;      FILING DATE:  19-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lambiris, Elias J.
;      REGISTRATION NUMBER:  33,728
;      REFERENCE/DOCKET NUMBER:  4006.204-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-867-0123
;      TELEFAX:  212-878-9655
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  295 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Aspergillus aculeatus
US-09-408-257-18


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-408-257-18   ..

 Align seg 1/1  to: US-09-408-257-18  from: 1  to: 295

     265 ProValGlyThrAlaAlaIleAla 272
         ||||||||||||||||||||||||
     114 CCTGTTGGCACCGCTGCTATTGCT 137

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-329-350-30
```

```
seq_documentation_block:
; Sequence 30, Application US/09329350
; Patent No. 6184019
;  GENERAL INFORMATION:
;    APPLICANT:  Miettinen-Oinonen, Arja
;    APPLICANT:  Londesborough, John
;    APPLICANT:  Vehmaanper , Jari
;    APPLICANT:  Haakana, Heli
;    APPLICANT:  M ntyl , Arja
;    APPLICANT:  Lantto, Raija
;    APPLICANT:  Elovainio, Minna
;    APPLICANT:  Joutsjoki, Vesa
;    APPLICANT:  Paloheimo, Marja
;    APPLICANT:  Suominen, Pirkko
;    TITLE OF INVENTION:  NOVEL CELLULASES, THE GENES ENCODING THEM AND
;    TITLE OF INVENTION:  USES THEREOF
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox P.L.L.C.
;      STREET:  1100 New York Avenue, N.W., Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/329,350
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/841,636
;      FILING DATE:  30-APR-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/005,335
;      FILING DATE:  17-OCT-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/007,926
;      FILING DATE:  04-DEC-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/020,840
;      FILING DATE:  28-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/732,181
;      FILING DATE:  16-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FI96/00550
;      FILING DATE:  17-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shea Jr., Timothy
;      REGISTRATION NUMBER:  41,306
;      REFERENCE/DOCKET NUMBER:  1716.0510006/MAC/TJS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)371-2600
;      TELEFAX:  (202)371-2540
;  INFORMATION FOR SEQ ID NO:  30:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  936 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Melanocarpus albomyces
;      STRAIN:  ALKO4237
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  33..115
;      OTHER INFORMATION:  /codon_start= 33
;      OTHER INFORMATION:  /product= "20K-cellulase"
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  187..435
;      OTHER INFORMATION:  /product= "20K-cellulase"
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  506..881
;      OTHER INFORMATION:  /product= "20K-cellulase"
US-09-329-350-30


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-329-350-30/rev   ..

 Align seg 1/1  to reverse of: US-09-329-350-30  from: 1  to: 936

     800 AlaGlyGlyLeuGluAspGlyGlu 807
         ||||||||||||||||||||||||
     875 GCTGGGGGCCTTGAAGACGGCGAA 852

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-482-918-60

seq_documentation_block:
; Sequence 60, Application US/08482918
; Patent No. 6207417
;  GENERAL INFORMATION:
;    APPLICANT:  Zsebo, Krisztina M.
;    APPLICANT:  Bosselman, Robert A.
;    APPLICANT:  Suggs, Sidney V.
;    APPLICANT:  Martin, Francis H.
;    TITLE OF INVENTION:  Stem Cell Factor
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/482,918
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  01017/33005
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  184..1002
;    FEATURE:
```

```
;       NAME/KEY:  mat_peptide
;       LOCATION:  259..1002
US-08-482-918-60


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-08-482-918-60   ..

 Align seg 1/1  to: US-08-482-918-60  from: 1  to: 1404

      64 AlaAlaGlySerArgAspSerGly 71
         ||||||||||||||||||||||||
      25 GCTGCGGGAAGCAGGGACAGTGGA 48

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-224-681-60

seq_documentation_block:
; Sequence 60, Application US/09224681
; Patent No. 6207454
;  GENERAL INFORMATION:
;    APPLICANT:  Zsebo, Krisztina M.
;    APPLICANT:  Bosselman, Robert A.
;    APPLICANT:  Suggs, Sidney V.
;    APPLICANT:  Martin, Francis H.
;    TITLE OF INVENTION:  Method for Enhancing the Efficiency of Gene
;    TITLE OF INVENTION:  Transfer with Stem Cell Factor (SCF), Polypeptide
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/224,681
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/005,893
;      FILING DATE:  12-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/449,653
;      FILING DATE:  24-MAY-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/982,255
;      FILING DATE:  25-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/589,701
;      FILING DATE:  01-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/573,616
;      FILING DATE:  24-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/537,198
;      FILING DATE:  11-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/422,383
;      FILING DATE:  16-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  01017/35199
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  184..1002
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  259..1002
US-09-224-681-60


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-224-681-60   ..

 Align seg 1/1  to: US-09-224-681-60  from: 1  to: 1404

      64 AlaAlaGlySerArgAspSerGly 71
         ||||||||||||||||||||||||
      25 GCTGCGGGAAGCAGGGACAGTGGA 48

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-336-728A-60

seq_documentation_block:
; Sequence 60, Application US/08336728A
; Patent No. 6207802
;  GENERAL INFORMATION:
;    APPLICANT:  Zsebo, Krisztina M.
;    APPLICANT:  Bosselman, Robert A.
;    APPLICANT:  Suggs, Sidney V.
;    APPLICANT:  Martin, Francis H.
;    TITLE OF INVENTION:  Stem Cell Factor
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/336,728A
;      FILING DATE:  09-NOV-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/982,255
;      FILING DATE:  25-NOV-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/589,701
;      FILING DATE:  01-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/573,616
;      FILING DATE:  24-AUG-1990
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/537,198
;      FILING DATE:  11-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/422,383
;      FILING DATE:  16-OCT-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  01017/32956
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  60:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1404 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  184..1002
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  259..1002
US-08-336-728A-60


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-08-336-728A-60   ..

 Align seg 1/1  to: US-08-336-728A-60  from: 1  to: 1404

      64 AlaAlaGlySerArgAspSerGly 71
         |||||||||||||||||||||||||
      25 GCTGCGGGAAGCAGGGACAGTGGA 48

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-035-706-1

seq_documentation_block:
; Sequence 1, Application US/09035706
; Patent No. 6001622
;  GENERAL INFORMATION:
;    APPLICANT:  Dedhar, Shoukat
;    APPLICANT:  Hannigan, Greg
;    TITLE OF INVENTION:  Integrin-Linked Kinase and
;    TITLE OF INVENTION:  its Uses
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Avenue, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/035,706
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  KIN-2CIP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3400
;      TELEFAX:  650 327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1789 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-035-706-1


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-09-035-706-1   ..

 Align seg 1/1  to: US-09-035-706-1  from: 1  to: 1789

     223 HisLeuProHisGlyValSerGln 230
         ||||||||||||||||||||||||
    1700 CACTTGCCACATGGTGTCTCCCAA 1723

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-955-841-1

seq_documentation_block:
; Sequence 1, Application US/08955841
; Patent No. 6013782
;  GENERAL INFORMATION:
;    APPLICANT:  Dedhar, Shoukat
;    APPLICANT:  Hannigan, Greg
;    TITLE OF INVENTION:  Integrin-Linked Kinase and
;    TITLE OF INVENTION:  its Uses
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Avenue, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/955,841
;      FILING DATE:
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  KIN-2CIP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3400
;      TELEFAX:  650 327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  1789 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-955-841-1


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-08-955-841-1    ..

 Align seg 1/1  to: US-08-955-841-1  from: 1  to: 1789

     223 HisLeuProHisGlyValSerGln 230
         ||||||||||||||||||||||||
    1700 CACTTGCCACATGGTGTCTCCCAA 1723

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-669-536-1

seq_documentation_block:
; Sequence 1, Application US/08669536
; Patent No. 5910444
;  GENERAL INFORMATION:
;     APPLICANT:  MASUTA, CHIKARA
;     APPLICANT:  UEHARA, KYOKO
;     APPLICANT:  TANAKA, HIDEO
;     APPLICANT:  KUWATA, SHIGERU
;     TITLE OF INVENTION:  ORGANISMS IN WHICH THE EXPRESSION OF
;     TITLE OF INVENTION:  S-ADENOSYLHOMOCYSTEINE HYDROLASE GENE IS INHIBITED
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  BIRCH, STEWART, KOLASCH AND BIRCH
;       STREET:  PO BOX 747
;       CITY:  FALLS CHURCH
;       STATE:  VA
;       COUNTRY:  USA
;       ZIP:  22040-0747
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/669,536
;       FILING DATE:
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MURPHY JR, GERALD M
;       REGISTRATION NUMBER:  28,977
;       REFERENCE/DOCKET NUMBER:  1254-128
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (703) 205-8000
;       TELEFAX:  (703) 205-8050
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1812 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
US-08-669-536-1


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-08-669-536-1    ..

 Align seg 1/1  to: US-08-669-536-1  from: 1  to: 1812

     679 AspAlaThrLeuLeuIleHisGlu 686
         ||||||||||||||||||||||||
     487 GATGCTACACTCTTGATTCATGAG 510

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-929-580B-1

seq_documentation_block:
; Sequence 1, Application US/07929580B
; Patent No. 5426181
;  GENERAL INFORMATION:
;     APPLICANT:  Lee, Tae Ho
;     APPLICANT:  Lee, Gene W.
;     APPLICANT:  Vilcek, Jan
;     TITLE OF INVENTION:  Cytokine-Induced Protein, TSG-14,
;     TITLE OF INVENTION:  DNA Coding Therefor and Uses Thereof
;     NUMBER OF SEQUENCES:  5
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Browdy and Neimark
;       STREET:  419 Seventh Street, NW
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  US
;       ZIP:  20004
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.24
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/929,580B
;       FILING DATE:  19920814
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/640,492
;       FILING DATE:  14-JAN-1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Townsend, Guy Kevin
;       REGISTRATION NUMBER:  34,033
;       REFERENCE/DOCKET NUMBER:  LEE25\VILCEK=2A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 628-5197
;       TELEFAX:  (202) 737-3528
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1836 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  73..606
US-07-929-580B-1


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-07-929-580B-1/rev    ..

 Align seg 1/1  to reverse of: US-07-929-580B-1  from: 1  to: 1836

     624 LeuIleSerSerLeuLeuArgThr 631
         ||||||||||||||||||||||||
    1116 CTTATCTCTTCATTGCTAAGAACA 1093
```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-929-580B-4

seq_documentation_block:
; Sequence 4, Application US/07929580B
; Patent No. 5426181
;  GENERAL INFORMATION:
;    APPLICANT:  Lee, Tae Ho
;    APPLICANT:  Lee, Gene W.
;    APPLICANT:  Vilcek, Jan
;    TITLE OF INVENTION:  Cytokine-Induced Protein, TSG-14,
;    TITLE OF INVENTION:  DNA Coding Therefor and Uses Thereof
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, NW
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  US
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/929,580B
;      FILING DATE:  19920814
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/640,492
;      FILING DATE:  14-JAN-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Townsend, Guy Kevin
;      REGISTRATION NUMBER:  34,033
;      REFERENCE/DOCKET NUMBER:  LEE25\VILCEK=2A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 628-5197
;      TELEFAX:  (202) 737-3528
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1839 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  74..1216
US-07-929-580B-4


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x US-07-929-580B-4/rev   ..

 Align seq 1/1  to reverse of: US-07-929-580B-4  from: 1  to: 1839

     624 LeuIleSerSerLeuLeuArgThr 631
         |||||||||||||||||||||||
    1119 CTTATCTCTTCATTGCTAAGAACA 1096
```

```
OM of: US-09-988-687-2 to: EST:*   out_format : pfs

Date: Mar 29, 2002  6:06 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=x1h
 -Q=/cgn2_1/USPTO_spool/US09988687/runat_29032002_092148_558/app_query.fasta_1.895
 -DB=EST -QFMT=fastap -SUFFIX=olip2n.rst -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09988687_@CGN1_1_4602 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-988-687-2
Query length: 826
Database: EST:*
Database sequences: 11351937
Database length: 1077921985
Search time (sec): 1176.950000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_est1:BE260495       +   225.00 3949.63   7.8e-211  676   ! BE260495 601150702F1 NIH_MGC_19
gb_est1:AU126037       +   213.00 3737.19   5.3e-199  712   ! AU126037 AU126037 NT2RM4 Homo s
gb_est1:AU141334       +   203.00 3561.04   3.5e-189  666   ! AU141334 AU141334 THYRO1 Homo s
gb_est1:AU143668       +   203.00 3561.00   3.5e-189  670   ! AU143668 AU143668 Y79AA1 Homo s
gb_est2:BF969043       +   199.00 3487.55   4.3e-185  944   ! BF969043 602269966F1 NIH_MGC_84
gb_est1:BE383336       +   197.00 3455.06   2.8e-183  664   ! BE383336 601298249F1 NIH_MGC_19
gb_est1:BE382353       +   196.00 3437.06   2.8e-182  692   ! BE382353 601298656F1 NIH_MGC_19
gb_est2:BF797306       +   194.00 3399.07   3.6e-180  960   ! BF797306 602256926F1 NIH_MGC_85
gb_est2:BG396395       +   190.00 3329.60   2.7e-176  827   ! BG396395 602459323F1 NIH_MGC_16
gb_est2:BG719016       +   186.00 3259.88   2.1e-172  735   ! BG719016 602699144F1 NIH_MGC_97
gb_est2:BG386090       +   186.00 3258.07   2.6e-172  920   ! BG386090 602455264F1 NIH_MGC_15
gb_est2:BG386348       +   185.00 3239.63   2.8e-171 1012   ! BG386348 602455550F1 NIH_MGC_1
gb_est2:BF434169       +   184.00 3226.85   1.4e-170  553   ! BF434169 7o99f04.x1 NCI_CGAP_Ov
gb_est2:BG519751       +   184.00 3223.05   2.3e-170  884   ! BG519751 602578902F1 NIH_MGC_19
gb_est2:BG756043       +   184.00 3223.00   2.3e-170  890   ! BG756043 602716533F1 NIH_MGC_48
gb_est2:BG746363       +   178.00 3117.05   1.9e-164  883   ! BG746363 602703625F1 NIH_MGC_15
gb_est2:BG335963       +   177.00 3102.32   1.2e-163  614   ! BG335963 602404708F1 NIH_MGC_21
gb_est2:BF196304       +   175.00 3068.24   9.7e-162  526   ! BF196304 7n68a10.x1 NCI_CGAP_Ov
gb_est1:AU138595       +   175.00 3065.97   1.3e-161  696   ! AU138595 AU138595 PLACE1 Homo s
gb_est1:AU124671       +   172.00 3012.24   1.3e-158  762   ! AU124671 AU124671 NT2RM4 Homo s
gb_est2:BF057321       +   168.00 2944.87   7.2e-155  506   ! BF057321 7k19c01.x1 NCI_CGAP_Ov
gb_est2:BG822529       +   165.00 2888.25   1.0e-151  791   ! BG822529 602725613F1 NIH_MGC_15
gb_est2:BF794577       +   165.00 2886.75   1.3e-151  953   ! BF794577 602255843F1 NIH_MGC_85
gb_est1:AU138795       +   159.00 2783.75   6.8e-146  657   ! AU138795 AU138795 PLACE1 Homo s
gb_est2:BE795820       +   159.00 2782.56   8.0e-146  761   ! BE795820 601590856F1 NIH_MGC_7
gb_est1:BE304720       +   158.00 2766.65   6.1e-145  612   ! BE304720 601106236F1 NIH_MGC_15
gb_est1:AW206103       +   157.00 2750.06   5.1e-144  536   ! AW206103 UI-H-BI1-afx-f-12-0-UI
gb_est2:BE794311       +   154.00 2695.03   6.0e-141  688   ! BE794311 601591442F1 NIH_MGC_7
gb_est1:BE619259       +   154.00 2693.58   7.2e-141  823   ! BE619259 601473130F1 NIH_MGC_68
gb_est1:AA243700       +   152.00 2661.79   4.2e-139  531   ! AA243700 zr68g08.s1 Soares_NhHM
gb_est2:BG766403       +   151.00 2641.10   6.0e-138  771   ! BG766403 602739146F1 NIH_MGC_49
gb_est1:AI200296       +   150.00 2627.77   3.3e-137  451   ! AI200296 qf86b12.x1 Soares_feta
gb_est1:BE260626       +   149.00 2604.18   6.9e-136  938   ! BE260626 601146116F1 NIH_MGC_19
gb_est2:BF433563       +   146.00 2556.30   3.2e-133  498   ! BF433563 7q55b07.x1 NCI_CGAP_Lu
gb_est1:AA455121       +   145.00 2538.71   3.0e-132  493   ! AA455121 zx78c04.s1 Soares ovar
gb_est2:BE892893       +   145.00 2534.90   5.0e-132  790   ! BE892893 601435738F1 NIH_MGC_72
gb_est1:BE250309       +   144.00 2519.81   3.4e-131  574   ! BE250309 600943455F1 NIH_MGC_17
gb_est2:BG751942       +   144.00 2517.54   4.6e-131  760   ! BG751942 602732110F1 NIH_MGC_43
gb_est1:AI468143       +   140.00 2451.65   2.2e-127  421   ! AI468143 tf92g05.x1 NCI_CGAP
gb_est1:AW575677       +   136.00 2379.70   2.2e-123  493   ! AW575677 UI-HF-BM0-adl-b-07-
gb_est2:BG327066       +   135.00 2362.73   1.9e-122  452   ! BG327066 602426274F1 NIH_MGC
gb_est2:BF309436       +   135.00 2362.64   2.0e-122  457   ! BF309436 601892128F1 NIH_MGC
gb_est1:AW304130       +   134.00 2345.97   1.7e-121  404   ! AW304130 xs13e05.x1 NCI_CGAP
gb_est1:AA291670       +   133.00 2327.01   1.9e-120  474   ! AA291670 zt37d04.s1 Soares o
gb_est1:BE386924       +   131.00 2289.41   2.3e-118  627   ! BE386924 601274815F1 NIH_MGC

seq_name: gb_est1:BE260495

seq_documentation_block:
LOCUS       BE260495      676 bp    mRNA            EST       26-OCT-2000
DEFINITION  601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
            mRNA sequence.
ACCESSION   BE260495
VERSION     BE260495.1  GI:9131807
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 676)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM176  row: d  column: 01
            High quality sequence stop: 672.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503184"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      154 a    207 c    176 g    139 t
ORIGIN


alignment_scores:
           Quality:  225.00            Length:    225
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BE260495   ..

 Align seg 1/1  to: BE260495  from: 1  to: 676

     365 ThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuAr 381
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCG 50

     381 gSerHisLysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheP 398
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 CAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCC 100
```

```
    398 roLeuLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerVal 414
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    101 CCCTGCTCACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTG 150

    415 ProMetValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgAr 431
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    151 CCCATGGTTCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAG 200

    431 gGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleV 448
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAG 250

    448 alGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArg 464
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    251 TTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 300

    465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrPr 481
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    301 AGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCC 350

    481 oGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    351 AGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAA 400

    498 snValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeu 514
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTG 450

    515 AspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGl 531
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    451 GACTGTGGTGAGGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCA 500

    531 nValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuH 548
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    501 GGTGGACAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGC 550

    548 isAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArg 564
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    551 ACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGC 600

    565 AlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeuValValAlaPr 581
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    601 GCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCC 650

    581 oAsnGlnLeuLysAlaTrpLeuGln 589
        |||||||||||||||||||||||||
    651 CAACCAGCTCAAAGCCTGGCTCCAG 675

seq_name: gb_est1:AU126037

seq_documentation_block:
LOCUS       AU126037      712 bp    mRNA            EST       23-OCT-2000
DEFINITION  AU126037 NT2RM4 Homo sapiens cDNA clone NT2RM4002610 5', mRNA
            sequence.
ACCESSION   AU126037
VERSION     AU126037.1  GI:10950753
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RM4002610"
                     /clone_lib="NT2RM4"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
BASE COUNT      173 a    198 c    204 g    134 t      3 others
ORIGIN


alignment_scores:
            Quality:  213.00             Length:   213
              Ratio:  1.000                Gaps:     0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AU126037   ..

 Align seg 1/1  to: AU126037  from: 1  to: 712

      1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     22 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 71

     17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
     72 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 121

     34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    122 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 171

     51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    172 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 221

     67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    222 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 271

     84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    272 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 321

    101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    322 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 371

    117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    372 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 421

    134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    422 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 471

    151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    472 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 521

    167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     522 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 571

     184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     572 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 621

     201 GluArgProLeuSerArgLeuSerProGluArgSerSer 213
         |||||||||||||||||||||||||||||||||||||
     622 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCA 660

seq_name: gb_est1:AU141334

seq_documentation_block:
LOCUS       AU141334      666 bp    mRNA            EST       25-OCT-2000
DEFINITION  AU141334 THYRO1 Homo sapiens cDNA clone THYRO1000421 5', mRNA
            sequence.
ACCESSION   AU141334
VERSION     AU141334.1  GI:11002855
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="THYRO1000421"
                     /clone_lib="THYRO1"
                     /tissue_type="thyroid gland"
                     /note="Vector: pME18SFL3"
BASE COUNT      158 a    186 c    190 g    129 t      3 others
ORIGIN


alignment_scores:
            Quality:  203.00             Length:    203
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AU141334   ..

 Align seg 1/1  to: AU141334  from: 1  to: 666

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      29 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 78

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      79 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 128

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     129 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 178

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     179 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 228

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     229 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 278

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     279 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 328

     101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     329 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 378

     117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     379 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 428

     134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     429 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 478

     151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     479 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 528

     167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     529 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 578

     184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     579 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 628

     201 GluArgPro 203
         |||||||||
     629 GAAAGGCCT 637

seq_name: gb_est1:AU143668

seq_documentation_block:
LOCUS       AU143668      670 bp    mRNA            EST       25-OCT-2000
DEFINITION  AU143668 Y79AA1 Homo sapiens cDNA clone Y79AA1002297 5', mRNA
            sequence.
ACCESSION   AU143668
VERSION     AU143668.1  GI:11005189
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .670
```

```
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="Y79AA1002297"
                    /clone_lib="Y79AA1"
                    /cell_type="retinoblastoma"
                    /cell_line="Y79"
                    /note="Vector: pME18SFL3"
BASE COUNT      157 a    190 c    193 g    127 t      3 others
ORIGIN


alignment_scores:
            Quality:  203.00             Length:    203
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AU143668   ..

 Align seg 1/1  to: AU143668  from: 1  to: 670

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      31 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 80

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      81 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 130

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     131 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 180

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     181 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 230

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     231 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 280

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     281 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 330

     101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     331 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 380

     117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     381 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 430

     134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     431 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 480

     151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     481 ATCAAAATATTTTCTGGCCCATTGAAAGGAATAGAACTGGCCGTGCGGCC 530

     167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     531 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 580

     184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     581 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 630

     201 GluArgPro 203
         |||||||||
     631 GAAAGGCCT 639
```

```
seq_name: gb_est2:BF969043

seq_documentation_block:
LOCUS       BF969043      944 bp    mRNA            EST       22-JAN-2001
DEFINITION  602269966F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357985 5',
            mRNA sequence.
ACCESSION   BF969043
VERSION     BF969043.1  GI:12336258
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9995  row: d  column: 18
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .944
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4357985"
                     /clone_lib="NIH_MGC_84"
                     /tissue_type="adrenal cortex carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
                     primed. Average insert size 1.229 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      200 a    266 c    264 g    214 t
ORIGIN


alignment_scores:
            Quality:  199.00             Length:    199
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BF969043   ..

 Align seg 1/1  to: BF969043  from: 1  to: 944

     266 ValGlyThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLy 282
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAA 50

     282 sSerIleThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrP 299
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 AAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTC 100

     299 roProAspProGlyAlaAlaPheValValValGluCysProAspGluSer 315
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CTCCAGATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGC 150

     316 PheIleGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLy 332
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 TTCATTCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAA 200

     332 sAlaAspAlaProValAlaLeuValValHisMetAlaProAlaSerValL 349
         |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
    201 GGCAGATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGC 250

    349 euValAspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThr 365
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 TTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACC 300

    366 GlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSe 382
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 CAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAG 350

    382 rHisLysIleGlnThrGlnLeuAsnLeuIleHisProAspIleProL 399
        ||||||||||||||||||||||||||||||||||||||||||||||||
    351 CCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCC 400

    399 euLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerValPro 415
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 TGCTCACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCC 450

    416 MetValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGl 432
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 ATGGTTCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGA 500

    432 uTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValG 449
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 GTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTG 550

    449 luAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArg 464
        |||||||||||||||||||||||||||||||||||||||||||||||
    551 AGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 597

seq_name: gb_est1:BE383336

seq_documentation_block:
LOCUS       BE383336      664 bp    mRNA            EST       21-JUL-2000
DEFINITION  601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
            mRNA sequence.
ACCESSION   BE383336
VERSION     BE383336.1  GI:9328701
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM313  row: a  column: 13
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3628308"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."

BASE COUNT      154 a    199 c    174 g    137 t
ORIGIN


alignment_scores:
           Quality:  197.00            Length:    197
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BE383336  . ..

 Align seg 1/1  to: BE383336  from: 1  to: 664

    365 ThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuAr 381
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCG 50

    381 gSerHisLysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheP 398
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 CAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCC 100

    398 roLeuLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerVal 414
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 CCCTGCTCACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTG 150

    415 ProMetValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgAr 431
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 CCCATGGTTCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAG 200

    431 gGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleV 448
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAG 250

    448 alGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArg 464
        ||||||||||||||||||||||||||||||||||||||||||||||||
    251 TTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 300

    465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrPr 481
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 AGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCC 350

    481 oGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 AGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAA 400

    498 snValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeu 514
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 ATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTG 450

    515 AspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGl 531
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 GACTGTGGTGAGGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCA 500

    531 nValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuH 548
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 GGTGGACAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGC 550

    548 isAlaAspHisHisThrGlyLeuProSerIleLeuLeuGln 561
        |||||||||||||||||||||||||||||||||||||||
    551 ACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTGCAA 591

seq_name: gb_est1:BE382353

seq_documentation_block:
LOCUS       BE382353      692 bp    mRNA            EST       21-JUL-2000
DEFINITION  601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
            mRNA sequence.
ACCESSION   BE382353
VERSION     BE382353.1  GI:9327718
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM314  row: o  column: 13
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1. .692
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3629028"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      152 a    193 c    197 g    150 t
ORIGIN


alignment_scores:
            Quality:  196.00            Length:    196
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BE382353   ..

 Align seg 1/1  to: BE382353  from: 1  to: 692

     493 ProMetLysIleArgAsnValSerAlaThrLeuValAsnIleSerProAs 509
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CCGATGAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGA 52

     509 pThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysA 526
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 CACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTGTGCC 102

     526 rgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAlaAlaVal 542
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 GTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCTGCTGTG 152

     543 PheValSerHisLeuHisAlaAspHisHisThrGlyLeuProSerIleLe 559
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 TTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTT 202

     559 uLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProL 576
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     203 GCTGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTT 252

     576 euLeuValValAlaProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHis 592
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     253 TGCTGGTGGTTGCCCCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCAC 302

     593 AsnGlnCysGlnGluValLeuHisHisIleSerMetIleProAlaLysCy 609
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 AACCAGTGCCAGGAGGTCCTGCACCACATCAGTATGATTCCTGCCAAATG 352

     609 sLeuGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleS 626
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 CCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCAGTGGAAAGATTGATCA 402

     626 erSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuVal 642
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     403 GTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGTCTGGTG 452

     643 ArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTr 659
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     453 CGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTG 502

     659 pLysValValTyrSerGlyAspThrMetProCysGluAlaLeuValArgM 676
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     503 GAAAGTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGA 552

     676 etGlyLysAspAlaThrLeuLeuIleHisGluAlaThr 688
         ||||||||||||||||||||||||||||||||||||
     553 TGGGGAAAGATGCCACCCTCCTGATACATGAAGCCACC 590

seq_name: gb_est2:BF797306

seq_documentation_block:
LOCUS       BF797306      960 bp    mRNA            EST       12-JAN-2001
DEFINITION  602256926F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4340286 5',
            mRNA sequence.
ACCESSION   BF797306
VERSION     BF797306.1  GI:12102360
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 960)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9952  row: c  column: 07
            High quality sequence stop: 705.
FEATURES             Location/Qualifiers
     source          1. .960
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4340286"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      230 a    286 c    276 g    168 t
ORIGIN


alignment_scores:
            Quality:  194.00            Length:    194
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BF797306   ..
```

```
Align seg 1/1  to: BF797306  from: 1  to: 960

    455 AsnPheGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyPr 471
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      2 AACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGCGCAGGACGGCCC 51

    471 oAlaProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyT 488
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     52 AGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATCTTCCTTGGAA 101

    488 hrGlySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuVal 504
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    102 CAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTC 151

    505 AsnIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyThrPh 521
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    152 AACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACGTT 201

    521 eGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyT 538
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    202 TGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCA 251

    538 hrLeuAlaAlaValPheValSerHisLeuHisAlaAspHisHisThrGly 554
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    252 CCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGC 301

    555 LeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLy 571
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    302 TTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAA 351

    571 sProLeuHisProLeuLeuValValAlaProAsnGlnLeuLysAlaTrpL 588
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    352 GCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGCTCAAAGCCTGGC 401

    588 euGlnGlnTyrHisAsnGlnCysGlnGluValLeuHisHisIleSerMet 604
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    402 TCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCACCACATCAGTATG 451

    605 IleProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerProAlaVa 621
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    452 ATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCAGT 501

    621 lGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluGluPheG 638
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    502 GGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATCTGGAAGAGTTTC 551

    638 lnThrCysLeuValArgHisCysLysHisAla 648
        |||||||||||||||||||||||||||||||
    552 AGACCTGTCTGGTGCGGCACTGCAAGCATGCG 583

seq_name: gb_est2:BG396395

seq_documentation_block:
LOCUS       BG396395      827 bp    mRNA            EST       12-MAR-2001
DEFINITION  602459323F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4581700 5',
            mRNA sequence.
ACCESSION   BG396395
VERSION     BG396395.1  GI:13289941
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1302  row: n  column: 05
            High quality sequence stop: 779.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4581700"
                     /clone_lib="NIH_MGC_16"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      205 a    217 c    226 g    179 t
ORIGIN


alignment_scores:
            Quality:  190.00            Length:    204
              Ratio:  0.936               Gaps:      1
 Percent Similarity:  99.510   Percent Identity: 99.510

alignment_block:
 US-09-988-687-2 x BG396395   ..

 Align seg 1/1  to: BG396395  from: 1  to: 827

    237 SerSerLeuValValAlaPheIleCysLysLeuHisLeuLysArgGlyAs 253
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     24 TCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAA 73

    253 nPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyThrAlaA 270
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     74 CTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGGGACAGCTG 123

    270 laIleAlaProIleIleAlaAlaValLysAspGlyLysSerIleThrHis 286
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    124 CCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCATCACTCAT 173

    287 GluGlyArgGluIleLeuAlaGluGluLeuCysThrProProAspProGl 303
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    174 GAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCAGATCCTGG 223

    303 yAlaAlaPheValValValGluCysProAspGluSerPheIleGlnProI 320
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    224 TGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCA 273

    320 leCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaPro 336
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    274 TCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCC 323

    337 ValAlaLeuValValHisMetAlaProAlaSerValLeuValAspSerAr 353
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    324 GTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAG 373

    353 gTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValL 370
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    374 GTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCACTTGGTCC 423

    370 euAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGln 386
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    424 TGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAGATTCAA 473

    387 ThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPh 403
        |||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     474 ACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTT 523

     403 eArgCysLysLysGluGlyProThrLeuSerValProMet.ValGlnGly 419
         |||||||||||||||||||||||||||||||||||||||| ||||||||
     524 CCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGGTTCAGGGT 573

     420 GluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpGlnArgAs 436
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     574 GAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGA 623

     436 pAlaIleIle 439
         ||||||||||
     624 TGCCATTATT 633

seq_name: gb_est2:BG719016

seq_documentation_block:
LOCUS       BG719016      735 bp    mRNA            EST       08-MAY-2001
DEFINITION  602699144F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4831139 5',
            mRNA sequence.
ACCESSION   BG719016
VERSION     BG719016.1  GI:13998216
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10753  row: g  column: 12
            High quality sequence stop: 728.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4831139"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Organ: testis; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.2 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      178 a    202 c    214 g    141 t
ORIGIN


alignment_scores:
            Quality:  186.00            Length:    214
              Ratio:  0.877               Gaps:      2
 Percent Similarity:  99.065   Percent Identity: 99.065

alignment_block:
 US-09-988-687-2 x BG719016   ..

 Align seg 1/1  to: BG719016  from: 1  to: 735
```

```
       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      36 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 85

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      86 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 135

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     136 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 185

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     186 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 235

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     236 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 285

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     286 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 335

     101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     336 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 385

     117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     386 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 435

     134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     436 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 485

     151 IleLysIlePheSerGlyProLeu.LysGlyIleGluLeuAlaValArgP 167
         |||||||||||||||||||||||| ||||||||||||||||||||||||
     486 ATCAAAATATTTTCTGGTCCATTGAAAAGGAATAGAACTGGCTGTGCGGC 535

     167 roHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIle 183
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     536 CCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATC 585

     184 ProIleHisSerGluGlnArgArgGlyLysHisGlnProTrp.GlnSerP 200
         |||||||||||||||||||||||||||||||||||||| |||||||||
     586 CCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGGCAGAGTC 635

     200 roGluArgProLeuSerArgLeuSerProGluArgSer 212
         |||||||||||||||||||||||||||||||||||||
     636 CAGAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCT 673

seq_name: gb_est2:BG386090

seq_documentation_block:
LOCUS       BG386090      920 bp    mRNA            EST       12-MAR-2001
DEFINITION  602455264F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583679 5',
            mRNA sequence.
ACCESSION   BG386090
VERSION     BG386090.1  GI:13279536
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
```

```
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: NIH Intramural Sequencing Center
           Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1307  row: p  column: 16
          High quality sequence stop: 736.
FEATURES             Location/Qualifiers
     source          1. .920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583679"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      196 a    261 c    272 g    190 t      1 others
ORIGIN


alignment_scores:
           Quality:  186.00            Length:    186
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BG386090   ..

 Align seg 1/1  to: BG386090  from: 1  to: 920

     306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysGl 322
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      21 TTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTGTGA 70

     322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      71 GAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCT 120

     339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     121 TGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACCAG 170

     356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGl 372
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     171 CAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCACTTGGTCCTGAATGA 220

     372 uAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnL 389
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     221 GAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAGATTCAAACCCAGC 270

     389 euAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgCys 405
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     271 TCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCGCTGT 320

     406 LysLysGluGlyProThrLeuSerValProMetValGlnGlyGluCysLe 422
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     321 AAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCT 370

     422 uLeuLysTyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaIleI 439
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     371 CCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTA 420

     439 leThrCysAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsn 455
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     421 TTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAAC 470

     456 PheGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAl 472
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     471 TTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGC 520

     472 aProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThrG 489
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     521 CCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATCTTCCTTGGAACAG 570

     489 lySerAla 491
         ||||||||
     571 GGTCTGCA 578

seq_name: gb_est2:BG386348

seq_documentation_block:
LOCUS       BG386348     1012 bp    mRNA            EST       12-MAR-2001
DEFINITION  602455550F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583749 5',
            mRNA sequence.
ACCESSION   BG386348
VERSION     BG386348.1  GI:13279794
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1012)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: NIH Intramural Sequencing Center
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1308  row: c  column: 14
           High quality sequence stop: 675.
FEATURES             Location/Qualifiers
     source          1. .1012
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583749"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      246 a    273 c    281 g    210 t      2 others
ORIGIN


alignment_scores:
           Quality:  185.00            Length:    199
             Ratio:  0.934               Gaps:      1
 Percent Similarity:  99.497   Percent Identity: 99.497

alignment_block:
 US-09-988-687-2 x BG386348   ..

 Align seg 1/1  to: BG386348  from: 1  to: 1012

     300 ProAspProGlyAlaAlaPheValValValGluCysProAspGluSerPh 316
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       4 CCAGATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTT 53
```

```
     316 eIleGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysA 333
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      54 CATTCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGG 103

     333 laAspAlaProValAlaLeuValValHisMetAlaProAlaSerValLeu 349
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     104 CAGATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTT 153

     350 ValAspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGl 366
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     154 GTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCA 203

     366 n.HisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSer 382
         | |||||||||||||||||||||||||||||||||||||||||||||||||
     204 GTCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGC 253

     383 HisLysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLe 399
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     254 CACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCT 303

     399 uLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerValProM 416
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     304 GCTCACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCA 353

     416 etValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGlu 432
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     354 TGGTTCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAG 403

     433 TrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGl 449
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     404 TGGCAGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGA 453

     449 uAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgS 466
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     454 GGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGA 503

     466 erAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGlu 482
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     504 GTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAA 553

     483 IleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArg 497
         |||||||||||||||||||||||||||||||||||||||||||||
     554 ATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGA 598

seq_name: gb_est2:BF434169

seq_documentation_block:
LOCUS       BF434169      553 bp    mRNA            EST       29-NOV-2000
DEFINITION  7o99f04.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644670 3'
            similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION   BF434169
VERSION     BF434169.1  GI:11446441
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 507.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3644670"
                     /clone_lib="NCI_CGAP_Ov18"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCGACATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      143 a    150 c    153 g    107 t
ORIGIN


alignment_scores:
            Quality:  184.00              Length:    184
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BF434169   ..

 Align seg 1/1  to: BF434169  from: 1  to: 553

      31 ArgProArgLysAspProLeuArgHisLeuArgThrArgGluLysArgGl 47
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 CGGCCGCGCAAGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGG 50

      47 yProSerGlyCysSerGlyGlyProAsnThrValTyrLeuGlnValValA 64
         |||||||||||||||||||||||||||||||||||||||||||||||||||
      51 ACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGG 100

      64 laAlaGlySerArgAspSerGlyAlaAlaLeuTyrValPheSerGluPhe 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     101 CAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 150

      81 AsnArgTyrLeuPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGl 97
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     151 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGA 200

      97 uHisLysLeuLysValAlaArgLeuAspAsnIlePheLeuThrArgMetH 114
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GCACAAGTTAAAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGC 250

     114 isTrpSerAsnValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGlu 130
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     251 ACTGGTCTAATGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAA 300

     131 ThrGlyLeuProLysCysValLeuSerGlyProProGlnLeuGluLysTy 147
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     301 ACCGGGCTTCCAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATA 350

     147 rLeuGluAlaIleLysIlePheSerGlyProLeuLysGlyIleGluLeuA 164
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     351 CCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGG 400

     164 laValArgProHisSerAlaProGluTyrGluAspGluThrMetThrVal 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     401 CTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 450

     181 TyrGlnIleProIleHisSerGluGlnArgArgGlyLysHisGlnProTr 197
         |||||||||||||||||||||||||||||||||||||||||||||||||||
     451 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATG 500
```

```
     197 pGlnSerProGluArgProLeuSerArgLeuSerProGluArgSerSerA 214
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 GCAGAGTCCAGAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAG 550

     214 sp 214
         ||
     551 AC 552

seq_name: gb_est2:BG519751

seq_documentation_block:
LOCUS       BG519751      884 bp    mRNA            EST       02-APR-2001
DEFINITION  602578902F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
            mRNA sequence.
ACCESSION   BG519751
VERSION     BG519751.1  GI:13515513
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 884)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM176  row: d  column: 01
            High quality sequence stop: 859.
FEATURES             Location/Qualifiers
     source          1. .884
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503184"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      199 a    251 c    249 g    185 t
ORIGIN


alignment_scores:
           Quality:  184.00            Length:    198
             Ratio:  0.934               Gaps:      1
 Percent Similarity:  99.495   Percent Identity: 99.495

alignment_block:
 US-09-988-687-2 x BG519751   ..

 Align seg 1/1  to: BG519751  from: 1  to: 884

     439 IleThrCysAsnProGluGluPheIleValGluAlaLeuGlnLeuProAs 455
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     273 ATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAA 322

     455 nPheGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProA 472
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     323 CTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAG 372

     472 laProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThr 488
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     373 CCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATCTTCCTTGGAACA 422

     489 GlySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuValAs 505
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     423 GGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTCAA 472

     505 nIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheG 522
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     473 CATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTG 522

     522 lyGlnLeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThr 538
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     523 GGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACC 572

     539 LeuAlaAlaValPheValSerHisLeuHisAlaAspHisHisThrGlyLe 555
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     573 CTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTT 622

     555 uProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysP 572
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     623 GCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGC 672

     572 roLeuHisProLeuLeuValValAlaProAsnGlnLeuLysAlaTrpLeu 588
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     673 CGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGCTCAAAGCCTGGCTC 722

     589 GlnGlnTyrHisAsnGlnCysGlnGluValLeuHisHisIle.SerMetI 605
         |||||||||||||||||||||||||||||||||||||| |||||||
     723 CAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCACCACATTCAGTATGA 772

     605 leProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerProAlaVal 621
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     773 TTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCAGTG 822

     622 GluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGlu 635
         |||||||||||||||||||||||||||||||||||||||||
     823 GAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAA 864

seq_name: gb_est2:BG756043

seq_documentation_block:
LOCUS       BG756043      890 bp    mRNA            EST       15-MAY-2001
DEFINITION  602716533F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4856360 5',
            mRNA sequence.
ACCESSION   BG756043
VERSION     BG756043.1  GI:14066696
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1707  row: b  column: 09
            High quality sequence stop: 878.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4856360"
```

```
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      209 a    247 c    275 g    159 t
ORIGIN


alignment_scores:
            Quality:  184.00             Length:    184
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x BG756043   ..

 Align seg 1/1  to: BG756043  from: 1  to: 890

     643 ArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTr 659
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       2 CGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTG 51

     659 pLysValValTyrSerGlyAspThrMetProCysGluAlaLeuValArgM 676
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      52 GAAAGTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGA 101

     676 etGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly 692
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     102 TGGGGAAAGATGCCACCCTCCTGATACATGAAGCCACCCTGGAAGATGGT 151

     693 LeuGluGluGluAlaValGluLysThrHisSerThrThrSerGlnAlaIl 709
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     152 TTGGAAGAGGAAGCAGTGGAAAAGACACACAGCACAACGTCCCAAGCCAT 201

     709 eSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsnHisPheS 726
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     202 CAGCGTGGGGATGCGGATGAACGCGGAGTTCATTATGCTGAACCACTTCA 251

     726 erGlnArgTyrAlaLysValProLeuPheSerProAsnPheSerGluLys 742
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     252 GCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGCGAGAAA 301

     743 ValGlyValAlaPheAspHisMetLysValCysPheGlyAspPheProTh 759
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     302 GTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGAGACTTTCCAAC 351

     759 rMetProLysLeuIleProProLeuLysAlaLeuPheAlaGlyAspIleG 776
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     352 AATGCCCAAGCTGATTCCCCCACTGAAAGCCCTGTTTGCTGGCGACATCG 401

     776 luGluMetGluGluArgArgGluLysArgGluLeuArgGlnValArgAla 792
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     402 AGGAGATGGAGGAGCGCAGGGAGAAGCGGGAGCTGCGGCAGGTGCGGGCG 451

     793 AlaLeuLeuSerArgGluLeuAlaGlyGlyLeuGluAspGlyGluProGl 809
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     452 GCCCTCCTGTCCAGGGAGCTGGCAGGCGGCCTGGAGGATGGGGAGCCTCA 501

     809 nGlnLysArgAlaHisThrGluGluProGlnAlaLysLysValArgAlaG 826
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     502 GCAGAAGCGGGCCCACACAGAGGAGCCACAGGCCAAGAAGGTCAGAGCCC 551

     826 ln 826
         ||
     552 AG 553
```

```
OM of: US-09-988-687-2 to: N_Geneseq_1101:*   out_format : pfs

Date: Mar 30, 2002 12:57 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09988687/runat_29032002_092149_615/app_query.fasta_1.895
 -DB=N_Geneseq_1101 -QFMT=fastap -SUFFIX=olip2n.rng -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09988687_@CGN1_1_269 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-988-687-2
Query length: 826
Database: N_Geneseq_1101:*
Database sequences: 930621
Database length: 428662619
Search time (sec): 118.770000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence          Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA58453 +  826.00 14886.69  0.0       2958
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA52810 +  624.00 11238.11  0.0       2478
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC76445 +  559.00 10063.43  0.0       2546
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH14250 +  430.00 7731.37   0.0       2976  !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH05835 +  121.00 2160.32   3.0e-112  584   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60207 +   82.00 1426.94   2.1e-71   26664
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ80231 +   78.00 1390.10   2.4e-69   238   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60390 +   15.00  249.40   8.1e-06   326   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC38171 +   10.00  156.40   1.23      464   !
/SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT23314 -    9.00  146.28   4.51      161   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC89621 +    9.00  125.64   63.72     2517  !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC30375 -    8.00  128.36   44.95     158   !
/SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT18675 +    8.00  126.01   60.75     216   !
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX41495 -    8.00  121.69   105.73    384   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC27730 +    8.00  120.93   116.59    425   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH83188 +    8.00  120.53   122.66    448   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC04558 +    8.00  120.41   124.50    455   !
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX20881 -    8.00  120.17   128.45    470   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC39669 -    8.00  119.69   136.61    501   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH09626 -    8.00  118.76   153.90    567   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH06196 +    8.00  118.75   154.16    568   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH82489 -    8.00  118.36   161.99    598   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF14046 +    8.00  117.68   176.84    655   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF14173 -    8.00  117.50   181.00    671   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA02584 -    8.00  117.11   190.09    706   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAF13618 -    8.00  116.65   201.74    751   !
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV28651 +    8.00  116.24   212.60    793   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH05641 +    8.00  116.22   213.11    795   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA09341 +    8.00  115.48   234.51    878   !
/SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:AAT66541 -    8.00  115.00   249.41    936   !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA87680 +    8.00  113.43   304.88    1153  !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA46491 -    8.00  112.66   336.40    1277  !
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX00619 -    8.00  112.01   365.79    1393  !
/SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:AAQ11542 +    8.00  111.95   368.57    1404  !
/SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT04890 +    8.00  111.95   368.57    1404  !
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX81675 -    8.00  111.95   368.57    1404  !
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA13714 +    8.00  111.95   368.57    1404  !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAS10461 +    8.00  111.95   368.57    1404  !
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH41344 +    8.00  111.95   368.57    1404
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAS04124 +    8.00  111.95   368.57    1404
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAS04224 +    8.00  111.95   368.57    1404
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH23901 +    8.00  111.95   368.57    1404
/SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF89104 +    8.00  111.95   368.57    1404
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC76846 +    8.00  111.72   379.44    1447
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC46817 +    8.00  111.48   391.30    1494

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA58453

seq_documentation_block:
ID   AAA58453 standard; cDNA; 2958 BP.
XX
AC   AAA58453;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..2531
FT                   /*tag= a
FT                   /product= "HPC2"
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Claim 3; Page 98-100; 157pp; English.
XX
CC   The present sequence is the coding sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;


alignment_scores:
            Quality:  826.00              Length:    826
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAA58453   ..

 Align seg 1/1  to: AAA58453  from: 1  to: 2958

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 100

 17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
    ||||||||||||||||||||||||||||||||||||||||||||||||||
101 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 150

 34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
    ||||||||||||||||||||||||||||||||||||||||||||||||||
151 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 200

 51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
201 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 250

 67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
251 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 300

 84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
301 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 350

101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
351 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 400

117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
401 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 450

134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
451 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAAATACCTCGAAGCA 500

151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
501 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 550

167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
551 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 600

184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
601 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 650

201 GluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluSe 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
651 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTC 700

217 rAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGlyV 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
701 GAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGG 750

234 alArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLys 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
751 TCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAG 800

251 ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
801 AGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGG 850

267 yThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSerI 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
851 GACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCA 900

284 leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
901 TCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 950

301 AspProGlyAlaAlaPheValValValGluCysProAspGluSerPheIl 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
951 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCAT 1000

317 eGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1001 TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAG 1050

334 spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1051 ATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTG 1100

351 AspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHi 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1101 GACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCA 1150

367 sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1151 CTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 1200

384 ysIleGlnThrGlnLeuAsnLeuIleHisProAspIleProLeuLeu 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1201 AGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1250

401 ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1251 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGT 1300

417 lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpG 434
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1301 TCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGC 1350

434 lnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAla 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1351 AGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCG 1400

451 LeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgSerAl 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1401 CTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGC 1450

467 aGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI 484
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1451 GCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCA 1500

484 lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1501 TCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1550

501 AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl 517
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1551 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGG 1600

517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1601 TGAGGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACA 1650

534 rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1651 GGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGAT 1700

551 HisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAl 567
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1701 CACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 1750

567 aSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnL 584
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1751 ATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGC 1800

584 euLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHis 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1801 TCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1850

601 HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSe 617
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1851 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTC 1900
```

```
   617 rSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspL 634
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  1901 CAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATT 1950

   634 euGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGly 650
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  1951 TGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGC 2000

   651 CysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2001 TGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC 2050

   667 rMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuI 684
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2051 CATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGA 2100

   684 leHisGluAlaThrLeuGluAspGlyLeuGluGluGluAlaValGluLys 700
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2101 TACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2150

   701 ThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAl 717
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2151 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGC 2200

   717 aGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProL 734
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2201 GGAGTTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCC 2250

   734 euPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMet 750
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2251 TCTTCAGCCCCAACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATG 2300

   751 LysValCysPheGlyAspPheProThrMetProLysLeuIleProProLe 767
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2301 AAGGTCTGCTTTGGAGACTTTCCAACAATGCCCAAGCTGATTCCCCCACT 2350

   767 uLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgArgGluL 784
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2351 GAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAGCGCAGGGAGA 2400

   784 ysArgGluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAla 800
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2401 AGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2450

   801 GlyGlyLeuGluAspGlyGluProGlnGlnLysArgAlaHisThrGluGl 817
       |||||||||||||||||||||||||||||||||||||||||||||||||||
  2451 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGA 2500

   817 uProGlnAlaLysLysValArgAlaGln 826
       |||||||||||||||||||||||||||
  2501 GCCACAGGCCAAGAAGGTCAGAGCCCAG 2528

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA52810

seq_documentation_block:
ID   AAA52810 standard; cDNA; 2478 BP.
XX
AC   AAA52810;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Human sulphatase G cDNA.
XX
KW   Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2478
FT                   /*tag=  a
FT                   /partial
FT                   /product= "hSG"
XX
PN   WO200034327-A1.
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999;   99WO-AU01092.
XX
PR   09-DEC-1998;   98AU-0007624.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Hopwood JJ,  Litjens T,  Hu RL;
XX
DR   WPI; 2000-431273/37.
DR   P-PSDB; AAY99850.
XX
PT   Novel isolated DNA sequence which encodes human sulfatase G or its
PT   fragment useful in gene therapy for treating patients suffering from
PT   sulfatase deficiency  -
XX
PS   Claim 2; Page 29-30; 33pp; English.
XX
CC   The present sequence encodes human sulphatase G (hSG). hSG is
CC   not a member of the well-characterised CTPSR sulphatase family.
CC   It belongs to a family showing sequence similarity to a sulphatase
CC   from the marine bacterium Alteromonas carrageenovora. The hSG gene
CC   contains 23 exons and is located at chromosome 17p11.2. The present
CC   sequence is clone lambda29.1 from a human testes cDNA library. It was
CC   isolated using human EST sequences with sequence similarity to the
CC   non-CTPSR family as a probe to screen the library. The cDNA insert was
CC   subcloned and the DNA sequence of both strands was determined. The
CC   sequence may be used to treat a patient suffering from hSG deficiency
CC   by replacing, repairing, or compensating for a DNA sequence within that
CC   patient's genome.
XX
SQ   Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other;


alignment_scores:
           Quality:  624.00            Length:    826
             Ratio:  0.757               Gaps:      0
 Percent Similarity:  99.758   Percent Identity: 99.758

alignment_block:
 US-09-988-687-2 x AAA52810   ..

 Align seg 1/1  to: AAA52810  from: 1  to: 2478

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 50

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 100

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 150

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 200

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 250

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 300

     101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 301 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 350

 117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 351 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 400

 134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 401 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 450

 151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 451 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 500

 167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 501 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 550

 184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 551 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 600

 201 GluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluSe 217
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTT 650

 217 rAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGlyV 234
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 651 GAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGG 700

 234 alArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLys 250
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 701 TCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAG 750

 251 ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl 267
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 751 AGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGG 800

 267 yThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSerI 284
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 801 GACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCA 850

 284 leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 851 TCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 900

 301 AspProGlyAlaAlaPheValValValGluCysProAspGluSerPheIl 317
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 901 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCAT 950

 317 eGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA 334
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 951 TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAG 1000

 334 spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal 350
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1001 ATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTG 1050

 351 AspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHi 367
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1051 GACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCA 1100

 367 sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL 384
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1101 CTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 1150

 384 ysIleGlnThrGlnLeuAsnLeuIleHisProAspIleProLeuLeu 400
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1151 AGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1200

 401 ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa 417
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

1201 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGT 1250

 417 lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpG 434
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1251 TCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGC 1300

 434 lnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAla 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1301 AGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCG 1350

 451 LeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgSerAl 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1351 CTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGC 1400

 467 aGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1401 GCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCA 1450

 484 lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1451 TCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1500

 501 AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1501 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGG 1550

 517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1551 TGAGGGCACGTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACA 1600

 534 rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp 550
     ||||||||||||||||||||  ||||||||||||||||||||||||||||
1601 GGGTCCTGGGCACCCTGGCTACTGTGTTTGTGTCCCACCTGCACGCAGAT 1650

 551 HisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAl 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1651 CACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 1700

 567 aSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnL 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1701 ATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGC 1750

 584 euLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHis 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1751 TCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1800

 601 HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1801 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTC 1850

 617 rSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspL 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1851 CAGTCCTGCAGTGGAAAGATTGATCAGTTCCCTGTTGCGAACATGTGATT 1900

 634 euGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGly 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1901 TGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGC 1950

 651 CysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1951 TGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC 2000

 667 rMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuI 684
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2001 CATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGA 2050

 684 leHisGluAlaThrLeuGluAspGlyLeuGluGluGluAlaValGluLys 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2051 TACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2100

 701 ThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAl 717
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2101 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGC 2150
```

```
     717 aGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProL 734
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2151 GGAGTTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCC 2200

     734 euPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMet 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2201 TCTTCAGCCCCAACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATG 2250

     751 LysValCysPheGlyAspProPheThrMetProLysLeuIleProProLe 767
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2251 AAGGTCTGCTTTGGAGACTTTCCAACAATGCCCAAGCTGATTCCCCCACT 2300

     767 uLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgArgGluL 784
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2301 GAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAGCGCAGGGAGA 2350

     784 ysArgGluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAla 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2351 AGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2400

     801 GlyGlyLeuGluAspGlyGluProGlnGlnLysArgAlaHisThrGluGl 817
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2401 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGA 2450

     817 uProGlnAlaLysLysValArgAlaGln 826
         |||||||||||||||||||||||||||
    2451 GCCACAGGCCAAGAAGGTCAGAGCCCAG 2478
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC76445

seq_documentation_block:

```
ID   AAC76445 standard; cDNA; 2546 BP.
XX
AC   AAC76445;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42236.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3179-3180; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;
```

alignment_scores:

```
            Quality:  559.00              Length:    573
              Ratio:  0.977                 Gaps:      1
 Percent Similarity:  99.825   Percent Identity: 99.825
```

alignment_block:

US-09-988-687-2 x AAC76445 ..

Align seg 1/1 to: AAC76445 from: 1 to: 2546

```
     255 LeuValLeuLysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIl 271
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     361 TTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGGGACAGCTGCCAT 410

     271 eAlaProIleIleAlaAlaValLysAspGlyLysSerIleThrHisGluG 288
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     411 CGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCATCACTCATGAAG 460

     288 lyArgGluIleLeuAlaGluGluLeuCysThrProProAspProGlyAla 304
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     461 GAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCAGATCCTGGTGCT 510

     305 AlaPheValValValGluCysProAspGluSerPheIleGlnProIleCy 321
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     511 GCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTG 560

     321 sGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValA 338
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     561 TGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGG 610

     338 laLeuValValHisMetAlaProAlaSerValLeuValAspSerArgTyr 354
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     611 CCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTAC 660

     355 GlnGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAs 371
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     661 CAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCACTTGGTCCTGAA 710

     371 nGluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrG 388
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 711 TGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAGATTCAAACCC 760

 388 lnLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArg 404
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 761 AGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCGC 810

 405 CysLysLysGluGlyProThrLeuSerValProMetValGlnGlyGluCy 421
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 811 TGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATG 860

 421 sLeuLeuLysTyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaI 438
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 861 CCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCA 910

 438 leIleThrCysAsnProGluGluPheIleValGluAlaLeuGlnLeuPro 454
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 911 TTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCC 960

 455 AsnPheGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyPr 471
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 961 AACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGCGCAGGACGGCCC 1010

 471 oAlaProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyT 488
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1011 AGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATCTTCCTTGGAA 1060

 488 hrGlySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuVal 504
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1061 CAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTC 1110

 505 AsnIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyThrPh 521
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1111 AACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATT 1160

 521 eGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyT 538
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1161 TGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCA 1210

 538 hrLeuAlaAlaValPheValSerHisLeuHisAlaAspHisHisThrGly 554
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1211 CCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGC 1260

 555 LeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLy 571
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1261 TTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAA 1310

 571 sProLeuHisProLeuLeuValValAlaProAsnGlnLeuLysAlaTrpL 588
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1311 GCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGCTCAAAGCCTGGC 1360

 588 euGlnGlnTyrHisAsnGlnCysGlnGluValLeuHisHisIleSerMet 604
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1361 TCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCACCACATCAGTATG 1410

 605 IleProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerProAlaVa 621
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1411 ATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCAGT 1460

 621 lGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluGluPheG 638
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1461 GGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTC 1510

 638 lnThrCysLeuValArgHisCysLysHisAlaPheGlyCysAlaLeuVal 654
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1511 AGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTG 1560

 655 HisThrSerGlyTrpLysValValTyrSerGlyAspThrMetProCysGl 671
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1561 CACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACACCATGCCCTGCGA 1610

 671 uAlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaT 688
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1611 GGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGATACATGAAGCCA 1660

 688 hrLeuGluAspGlyLeuGluGluGluAlaValGluLysThrHisSerThr 704
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1661 CCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAGACACACAGCACA 1710

 705 ThrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMe 721
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1711 ACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATTAT 1760

 721 tLeuAsnHisPheSerGlnArgTyrAlaLysValProLeuPheSerProA 738
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1761 GCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCA 1810

 738 snPheSerGluLysValGlyValAlaPheAspHisMetLysValCysPhe 754
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1811 ACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTT 1860

 755 GlyAspPheProThrMetProLysLeuIle.ProProLeuLysAlaLeuP 771
     |||||||||||||||||||||||||||||| ||||||||||||||||||||
1861 GGAGACTTTCCAACAATGCCCAAGCTGATTCCCCCCACTGAAAGCCCTGT 1910

 771 heAlaGlyAspIleGluGluMetGluGluArgArgGluLysArgGluLeu 787
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1911 TTGCTGGCGACATCGAGGAGATGGAGGAGCGCAGGGAGAAGCGGGAGCTG 1960

 788 ArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAlaGlyGlyLeuGl 804
     |||||||||||||||||||||||||||||||||||||||||||||||||||
1961 CGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCAGGCGGCCTGGA 2010

 804 uAspGlyGluProGlnGlnLysArgAlaHisThrGluGluProGlnAlaL 821
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2011 GGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGAGCCACAGGCCA 2060

 821 ysLysValArgAlaGln 826
     |||||||||||||||||
2061 AGAAGGTCAGAGCCCAG 2077
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH14250

seq_documentation_block:

```
ID   AAH14250 standard; cDNA; 2976 BP.
XX
AC   AAH14250;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11557.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
```

```
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11557; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2976 BP; 712 A; 807 C; 856 G; 601 T; 0 other;


alignment_scores:
           Quality:  430.00            Length:    753
             Ratio:  0.573               Gaps:      1
 Percent Similarity:  99.602   Percent Identity: 99.602

alignment_block:
 US-09-988-687-2 x AAH14250   ..

 Align seg 1/1  to: AAH14250  from: 1  to: 2976

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      75 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 124

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     125 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 174

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     175 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 224

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     225 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 274

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     275 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 324

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     325 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 374

     101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     375 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 424

     117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     425 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 474

     134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     475 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 524

     151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
         |||||||||||||||||||||||||||||||||||||||||||||||||
     525 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTATGCGGCC 574

     167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     575 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 624

     184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     625 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 674

     201 GluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluSe 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     675 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTC 724

     217 rAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGlyV 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     725 GAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGG 774

     234 alArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     775 TCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAG 824

     251 ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     825 AGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGG 874

     267 yThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSerI 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     875 GACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCA 924

     284 leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     925 TCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 974

     301 AspProGlyAlaAlaPheValValValGluCysProAspGluSerPheIl 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     975 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCAT 1024

     317 eGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1025 TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAG 1074

     334 spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1075 ATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTG 1124

     351 AspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHi 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1125 GACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCA 1174

     367 sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1175 CTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 1224

     384 ysIleGlnThrGlnLeuAsnLeuIleHisProAspIleProLeuLeu 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1225 AGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1274

     401 ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa 417
         ||||||||||||||||  ||||||||||||||||||||||||||||||||
    1275 ACCAGTTTCCGCTGT...AAGGAGGGCCCCACCCTCAGTGTGCCCATGGT 1321

     417 lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpG 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1322 TCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGC 1371

     434 lnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAla 450
```

```
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1372 AGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCG 1421

 451 LeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgSerAl 467
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1422 CTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGC 1471

 467 aGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI 484
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1472 GCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCA 1521

 484 lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer 500
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1522 TCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1571

 501 AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl 517
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1572 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGG 1621

 517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1622 TGAGGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACA 1671

 534 rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp 550
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1672 GGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGAT 1721

 551 HisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAl 567
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1722 CACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 1771

 567 aSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnL 584
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1772 ATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGC 1821

 584 euLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHis 600
     ||||||||||||||||||||||   ||||||||||||||||||||||||||
1822 TCAAAGCCTGGCTCCAGCAGTACTACAACCAGTGCCAGGAGGTCCTGCAC 1871

 601 HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSe 617
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1872 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTC 1921

 617 rSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspL 634
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1922 CAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATT 1971

 634 euGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGly 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1972 TGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGC 2021

 651 CysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2022 TGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC 2071

 667 rMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuI 684
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2072 CATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGA 2121

 684 leHisGluAlaThrLeuGluAspGlyLeuGluGluGluAlaValGluLys 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2122 TACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2171

 701 ThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAl 717
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2172 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGC 2221

 717 aGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProL 734
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
2222 GGAGTTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCC 2271

 734 euPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMet 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

     2272 TCTTCAGCCCCAACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATG 2321

      751 LysValCys 753
          ||||||||
     2322 AAGGTCTGC 2330
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAH05835

seq_documentation_block:

ID AAH05835 standard; cDNA; 584 BP.
XX
AC AAH05835;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:2670.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 2670; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 584 BP; 122 A; 166 C; 178 G; 115 T; 3 other;

```
alignment_scores:
            Quality:  121.00              Length:    121
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAH05835   ..

 Align seg 1/1  to: AAH05835  from: 1  to: 584

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      75 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 124

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     125 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 174

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     175 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 224

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     225 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 274

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     275 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 324

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     325 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 374

     101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     375 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 424

     117 nValGlyGlyLeu 121
         ||||||||||||
     425 TGTTGGGGGCTTA 437

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60207

seq_documentation_block:
ID   AAA60207 standard; DNA; 26664 BP.
XX
AC   AAA60207;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 genomic sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             910..26039
FT                   /*tag= a
FT                   /product= "HPC2"
FT                   /note= "this sequence contains introns"
FT                   /transl_except= (pos:23892..23895,aa:Glu)
FT   exon            910..1154
FT                   /*tag= b
FT                   /number= 1
FT   exon            1736..1786
FT                   /*tag= c
FT                   /number= 2
FT   exon            1925..1995
FT                   /*tag= d
FT                   /number= 3
FT   exon            3025..3089
FT                   /*tag= e
FT                   /number= 4
FT   exon            4361..4418
FT                   /*tag= f
FT                   /number= 5
FT   exon            5582..5650
FT                   /*tag= g
FT                   /number= 6
FT   exon            7075..7194
FT                   /*tag= h
FT                   /number= 7
FT   exon            8186..8244
FT                   /*tag= i
FT                   /number= 8
FT   exon            12878..12936
FT                   /*tag= j
FT                   /number= 9
FT   exon            13032..13104
FT                   /*tag= k
FT                   /number= 10
FT   exon            13756..13868
FT                   /*tag= l
FT                   /number= 11
FT   exon            15283..15378
FT                   /*tag= m
FT                   /number= 12
FT   exon            16278..16416
FT                   /*tag= n
FT                   /number= 13
FT   exon            16498..16583
FT                   /*tag= o
FT                   /number= 14
FT   exon            18583..18701
FT                   /*tag= p
FT                   /number= 15
FT   exon            20349..20445
FT                   /*tag= q
FT                   /number= 16
FT   exon            22172..22310
FT                   /*tag= r
FT                   /number= 17
FT   exon            22879..22917
FT                   /*tag= s
FT                   /number= 18
FT   exon            23045..23154
FT                   /*tag= t
FT                   /number= 19
FT   exon            23795..23895
FT                   /*tag= u
FT                   /number= 20
FT   exon            23973..24093
FT                   /*tag= v
FT                   /number= 21
FT   exon            24354..24432
FT                   /*tag= w
FT                   /number= 22
FT   exon            25026..25170
FT                   /*tag= x
FT                   /number= 23
FT   exon            25812..26036
FT                   /*tag= y
FT                   /number= 24
FT   polyA_signal    26447..26452
FT                   /*tag= z
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
```

```
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Claim 3; Page 108-122; 157pp; English.
XX
CC   The present sequence is the genomic sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;


alignment_scores:
           Quality:   82.00            Length:     82
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAA60207   ..

 Align seg 1/1  to: AAA60207  from: 1  to: 26664

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     910 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 959

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     960 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 1009

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1010 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 1059

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1060 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 1109

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArg 82
         |||||||||||||||||||||||||||||||||||||||||||||||
    1110 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGG 1155

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAZ80231

seq_documentation_block:
ID   AAZ80231 standard; cDNA; 238 BP.
XX
AC   AAZ80231;
XX
DT   07-APR-2000  (first entry)
XX
DE   Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
XX
KW   Human; gene expression product; diagnosis; tumour; colon cancer;
KW   colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW   cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW   hyperplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO9964576-A2.
XX
PD   16-DEC-1999.
XX
PF   09-JUN-1999;   99WO-IB01062.
XX
PR   10-JUN-1998;   98US-0088801.
XX
PA   (FARB ) BAYER CORP.
XX
PI   Endege WO,  Steinmann KE,  Astle JH,  Burgess CC,  Bushnell SE;
PI   Carroll E,  Catino TJ,  Derti A,  Ford DM,  Lewis ME,  Monahan JE;
PI   Schlegel R;
XX
DR   WPI; 2000-087220/07.
XX
PT   Novel nucleic acids, used to develop products for the diagnosis and
PT   treatment of disorders involving unwanted cell proliferation,
PT   particularly cancers, especially colon cancer -
XX
PS   Claim 15; Page 258; 469pp; English.
XX
CC   AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC   the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC   cDNA clones can be used to generate antisense oligonucleotides which
CC   can be used for antisense therapy. Methods and products from the present
CC   invention can be used for identifying and/or classifying cancerous cells
CC   present in a human tumour, particularly in solid tumours, e.g.
CC   carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC   can be used for developing agents for the diagnosis and treatment of
CC   disorders involving unwanted cell proliferation, such as neoplasia,
CC   dysplasia or hyperplasia.
XX
SQ   Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;


alignment_scores:
           Quality:   78.00            Length:     78
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAZ80231   ..

 Align seg 1/1  to: AAZ80231  from: 1  to: 238

      60 LeuGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrVa 76
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       3 CTGCAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGT 52

      76 lPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGlyValGlnA 93
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      53 CTTCTCCGAGTTCAACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGA 102

      93 rgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePhe 109
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     103 GACTCATGCAGGAGCACAAGTTAAAGGTTGCTCGCCTGGACAACATATTC 152

     110 LeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIleLe 126
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     153 CTGACACGAATGCACTGGTCTAATGTTGGGGGCTTAAGTGGAATGATTCT 202

     126 uThrLeuLysGluThrGlyLeuProLysCysVal 137
         |||||||||||||||||||||||||||||||||
     203 TACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTA 236

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAA60390

seq_documentation_block:
ID   AAA60390 standard; cDNA; 326 BP.
XX
AC   AAA60390;
```

```
XX
DT   07-DEC-2000  (first entry)
XX
DE   Murine prostate cancer predisposing gene HPC2 coding sequence.
XX
KW   Mouse; prostate cancer predisposing gene; HPC2; gene therapy;
KW   peptide therapy; drug design; ss.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             51..269
FT                   /*tag=  a
FT                   /product= "Mm.HPC2"
FT                   /partial
FT                   /note= "this sequence contains no termination codon"
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07230.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer  -
XX
PS   Disclosure; Page 151; 157pp; English.
XX
CC   The present sequence is the coding sequence of the murine prostate
CC   cancer predisposing gene Mm. HPC2, the human homologue of which is found
CC   on chromosome 17p. Some alleles of this gene cause a predisposition to
CC   cancer, particularly prostate cancer. This gene and its protein can be
CC   used in peptide and gene therapy for cancer patients, as well as being
CC   useful as diagnostic tools (both for cancer sufferers and those with a
CC   predisposition to the disease) and in the production of cancer drugs.
XX
SQ   Sequence 326 BP; 39 A; 104 C; 127 G; 56 T; 0 other;


alignment_scores:
            Quality:   15.00              Length:     15
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAA60390   ..

 Align seg 1/1  to: AAA60390  from: 1  to: 326

      34 LysAspProLeuArgHisLeuArgThrArgGluLysArgGlyPro 48
         ||||||||||||||||||||||||||||||||||||||||||||||
     132 AAGGACCCACTGCGACACCTGCGTACGCGGGAGAAGCGCGGCCCG 176

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC38171

seq_documentation_block:
ID   AAC38171 standard; DNA; 464 BP.
XX
AC   AAC38171;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 20026.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
```

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-988-687-2 x AAC38171 ..

Align seg 1/1 to: AAC38171 from: 1 to: 464

511 SerLeuLeuLeuAspCysGlyGluGlyThr 520
||||||||||||||||||||||||||||||
200 AGTCTTCTCCTAGATTGTGGTGAAGGAACC 229

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:AAT23314

seq_documentation_block:
ID AAT23314 standard; cDNA to mRNA; 161 BP.
XX
AC AAT23314;
XX
DT 02-SEP-1996 (first entry)
XX
DE Human gene signature HUMGS05130.
XX
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;

```
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 1339-1340; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in AAT19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
SQ   Sequence 161 BP; 30 A; 56 C; 45 G; 30 T; 0 other;


alignment_scores:
          Quality:    9.00            Length:      9
            Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAT23314/rev   ..

 Align seg 1/1  to reverse of: AAT23314  from: 1  to: 161

     569 LeuGlyLysProLeuHisProLeuLeu 577
         ||||||||||||||||||||||||||||
      52 CTGGGCAAACCGCTGCATCCGCTCCTC 26

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:AAC89621

seq_documentation_block:
ID   AAC89621 standard; DNA; 2517 BP.
XX
AC   AAC89621;
XX
DT   08-MAR-2001  (first entry)
XX
DE   S. cerevisiae YKR079C gene.
XX
KW   Yeast; germination; proliferation; essential gene; antifungal agent;
KW   insecticide; herbicide; anti-proliferation drug; cancer; psoriasis;
KW   restenosis; YKR081C; YFR003C; YGR277C; YGR278W; YKR071C; YKR079C;
KW   YKR083C; ds.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200071161-A1.
XX
PD   30-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13017.
XX
PR   21-MAY-1999;   99US-0315794.
PR   02-SEP-1999;   99US-0389341.
XX
PA   (ROSE-) ROSETTA INPHARMATICS INC.
XX
PI   Roberts CJ;
XX
DR   WPI; 2001-025092/03.
DR   P-PSDB; AAB49964.
XX
PT   Identifying antifungal compounds which target yeast essential genes
PT   comprises use of novel Saccharomyces cerevisiae essential genes
PT   YFR003C, YGR277C, YGR278W, YKR0701C, YKR079C or YKR083C or YKR081C  -
XX
PS   Example 5; Fig 29; 127pp; English.
XX
CC   The present invention provides methods of identifying antifungal agents
CC   using the coding and protein sequences of several yeast genes. These are
CC   essential for the germination and proliferation of Saccharomyces
CC   cerevisiae, and include YKR081C, YFR003C, YGR277C, YGR278W, YKR071C,
CC   YKR079C and YKR083C. The sequences can also be used to identify compounds
CC   for use as herbicides, insecticides and anti-proliferation drugs which
CC   can be used in the treatment of cancer, psoriasis and restenosis. This is
CC   because they can be used to identify plant, insect and human homologues
CC   of the yeast genes.
XX
SQ   Sequence 2517 BP; 881 A; 417 C; 496 G; 723 T; 0 other;


alignment_scores:
          Quality:    9.00            Length:      9
            Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAC89621   ..

 Align seg 1/1  to: AAC89621  from: 1  to: 2517

     682 LeuLeuIleHisGluAlaThrLeuGlu 690
         ||||||||||||||||||||||||||||
    2149 CTATTAATTCACGAAGCTACACTAGAA 2175

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC30375

seq_documentation_block:
ID   AAC30375 standard; cDNA; 158 BP.
XX
AC   AAC30375;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 34450.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
```

```
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 34450; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 158 BP; 20 A; 54 C; 47 G; 37 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAC30375/rev   ..

 Align seg 1/1  to reverse of: AAC30375  from: 1  to: 158

     470 GlyProAlaProAlaGluLysArg 477
         ||||||||||||||||||||||||
      72 GGTCCAGCTCCTGCTGAGAAGCGC 49

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:AAT18675

seq_documentation_block:
ID   AAT18675 standard; cDNA; 216 BP.
XX
AC   AAT18675;
XX
DT   03-JUL-1996  (first entry)
XX
DE   Human trophinin external domain 2 cDNA.
XX
KW   Trophinin; trophinin-assisting protein; tastin; bystin; lastin;
KW   embryo implantation; infertility; cell adhesion; therapy; diagnosis;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   repeat_unit     1..30
FT                   /*tag=  a
FT   repeat_unit     31..60
FT                   /*tag=  b
FT   repeat_unit     61..90
FT                   /*tag=  c
FT   repeat_unit     91..120
FT                   /*tag=  d
FT   repeat_unit     121..150
FT                   /*tag=  e
FT   repeat_unit     151..180
FT                   /*tag=  f
FT   repeat_unit     181..210
FT                   /*tag=  g
XX
PN   WO9610414-A1.
XX
PD   11-APR-1996.
XX
PF   04-OCT-1995;   95WO-US13259.
XX
PR   12-MAY-1995;   95US-0439818.
PR   04-OCT-1994;   94US-0317522.
XX
PA   (LJOL-) LA JOLLA CANCER RES FOUND.
XX
PI   Fukuda MN;
XX
DR   WPI; 1996-209192/21.
DR   P-PSDB; AAR94898.
XX
PT   Mammalian trophinin and trophinin-assisting protein - used in
PT   inhibiting or enhancing embryo implantation, diagnosis of
PT   infertility and treatment of cancer
XX
PS   Claim 21; Fig 3; 106pp; English.
XX
CC   3 Fragments (AAT18674-76) of a human trophinin cDNA clone (AAT18673)
CC   code for active, exposed cell surface domains (AAR94897-99,
CC   respectively), of trophinin.  These domains contain regions of
CC   hydrophilic decapeptide repeats.  Protein secondary structure
CC   algorithms predict that the decapeptide repeats conform to a
CC   repeated beta-turn structure, which may be involved in homophilic
CC   adhesion.
XX
SQ   Sequence 216 BP; 36 A; 66 C; 54 G; 60 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAT18675   ..

 Align seg 1/1  to: AAT18675  from: 1  to: 216

     343 MetAlaProAlaSerValLeuVal 350
         ||||||||||||||||||||||||
      62 ATGGCACCAGCCTCTGTTTTGGTG 85

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:AAX41495

seq_documentation_block:
ID   AAX41495 standard; cDNA; 384 BP.
XX
AC   AAX41495;
XX
DT   22-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO: 154 from WO 9906553.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; antiinflammatory; tumour inhibition; antitumour; ds.
```

```
XX
OS   Homo sapiens.
XX
PN   WO9906553-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01237.
XX
PR   01-AUG-1997;   97US-0905051.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153783/13.
DR   P-PSDB; AAY12637.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from umbilical cord, lymph ganglia,
PT   lymphocytes and placental tissue
XX
PS   Claim 1; Page 271; 411pp; English.
XX
CC   AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12521 to
CC   AAY12668, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, antiinflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 384 BP; 76 A; 112 C; 105 G; 85 T; 6 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAX41495/rev   ..

 Align seg 1/1  to reverse of: AAX41495  from: 1  to: 384

       8 LeuArgSerAlaAlaGlyArgThr 15
         ||||||||||||||||||||||||
     358 TTGAGGTCAGCCGCAGGAAGAACT 335

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:AAC27730

seq_documentation_block:
ID   AAC27730 standard; cDNA; 425 BP.
XX
AC   AAC27730;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 31805.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 31805; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 425 BP; 127 A; 66 C; 115 G; 108 T; 9 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AAC27730   ..

 Align seg 1/1  to: AAC27730  from: 1  to: 425

     733 ProLeuPheSerProAsnPheSer 740
         ||||||||||||||||||||||||
     323 CCGCTTTTCTCGCCAAACTTCAGT 346
```

```
OM of: US-09-988-687-2 to: GenEmbl:*   out_format : pfs

Date: Mar 29, 2002  6:33 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09988687/runat_29032002_092149_570/app_query.fasta_1.895
 -DB=GenEmbl -QFMT=fastap -SUFFIX=olip2n.rge -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09988687_@CGN1_1_7060 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-988-687-2
Query length: 826
Database: GenEmbl:*
Database sequences: 1472140
Database length: -341344837
Search time (sec): 1563.220000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_pr:AF304370          +  826.00 16166.98  0.0       2908  ! AF304370 Homo sapiens putativ
gb_pr:BC001939          +  725.00 14184.43  0.0       2997  ! BC001939 Homo sapiens, putati
gb_pr:BC004158          +  725.00 14184.41  0.0       3006  ! BC004158 Homo sapiens, putati
gb_pr:AK001392          +  430.00 8394.51   0.0       2976  ! AK001392 Homo sapiens cDNA FLJ
gb_pr:AF308698          +  254.00 4940.31   6.7e-267  2908  ! AF308698 Pan troglodytes ELAC2
gb_pr:AF308694          +  157.00 3036.53   7.4e-161  2893  ! AF308694 Gorilla gorilla ELAC2
gb_pr:AC005277          +   82.00 1538.16   2.1e-77   118788  ! AC005277 Homo sapiens chromo
gb_ro:AF308696          +   51.00  956.52   5.3e-45   2712  ! AF308696 Mus musculus ELAC2 mR
gb_pr:AF304371S2        +   47.00  887.22   3.8e-41   740  ! AF304369 Homo sapiens putative
gb_pr:AF304371S1        +   40.00  750.02   1.7e-33   721  ! AF304371 Homo sapiens putative
gb_ro:AF348157          +   26.00  447.79   1.1e-16   34593  ! AF348157 Mus musculus putativ
gb_pl:SPAC1D4           +   11.00  154.12   2.62      31201  ! Z69239 S.pombe chromosome I c
gb_htg:AC036490         +   10.00  159.78   1.27      881  ! AC036490 Giardia intestinalis c
gb_htg:AC031419         +   10.00  159.34   1.34      938  ! AC031419 Giardia intestinalis c
gb_pl:AC022354          -   10.00  131.69   46.47     46335  ! AC022354 Arabidopsis thaliana
gb_pr:AL158046          +   10.00  124.99   109.77    119265  ! AL158046 Human DNA sequence
gb_pl:AP001550          +   10.00  123.69   129.64    143209  ! AP001550 Oryza sativa genomi
gb_pr:AC055736          +   10.00  123.32   135.96    150902  ! AC055736 Homo sapiens 12 BAC
gb_ro:AC003694          +   10.00  121.82   164.69    186314  ! AC003694 Mus musculus chromo
gb_ba:RME603644         +   10.00  118.10   265.49    315000  ! AL603644 Rhizobium meliloti
gb_sts:G26261           -    9.00  150.71   4.06      199  ! G26261 human STS TIGR-A002J17,
gb_htg:AC029747         -    9.00  141.02   14.04     780  ! AC029747 Giardia intestinalis c
gb_htg:AC032553         -    9.00  140.84   14.37     800  ! AC032553 Giardia intestinalis c
gb_htg:AC088618         +    9.00  140.84   14.37     800  ! AC088618 Giardia intestinalis c
gb_htg:AC082071         +    9.00  140.80   14.45     805  ! AC082071 Giardia intestinalis c
gb_htg:AC046748         +    9.00  140.48   15.05     842  ! AC046748 Giardia intestinalis c
gb_htg:AC063132         -    9.00  140.01   15.99     900  ! AC063132 Giardia intestinalis c
gb_htg:AC030275         +    9.00  139.94   16.14     909  ! AC030275 Giardia intestinalis c
gb_htg:AC040707         -    9.00  139.88   16.25     916  ! AC040707 Giardia intestinalis c
gb_htg:AC070405         +    9.00  139.86   16.30     919  ! AC070405 Giardia intestinalis c
gb_htg:AC033678         +    9.00  139.85   16.32     920  ! AC033678 Giardia intestinalis c
gb_htg:AC037172         -    9.00  139.83   16.37     923  ! AC037172 Giardia intestinalis c
gb_htg:AC040050         -    9.00  139.77   16.48     930  ! AC040050 Giardia intestinalis c
gb_sts:CNS06JSP         +    9.00  139.72   16.59     937  ! AL401999 T7 end of clone XAS0AA
gb_htg:AC051575         -    9.00  139.70   16.64     940  ! AC051575 Giardia intestinalis c
gb_sts:CNS06HXX         +    9.00  139.62   16.82     951  ! AL399595 T3 end of clone AS0AAC
gb_htg:AC084975         -    9.00  139.51   17.04     965  ! AC084975 Giardia intestinalis c
gb_htg:AC064766         +    9.00  138.89   18.45     1053  ! AC064766 Giardia intestinalis

gb_pr:BC007619          -    9.00  138.10   20.43     1178  ! BC007619 Homo sapiens, clon
gb_htg:AC049359         +    9.00  138.01   20.67     1193  ! AC049359 Giardia intestinal
gb_ba:AF038408          +    9.00  137.87   21.04     1217  ! AF038408 Streptoalloteichus
gb_pr:BC008146          -    9.00  136.30   25.71     1517  ! BC008146 Homo sapiens, clon
gb_pr:AF053070          -    9.00  136.08   26.47     1566  ! AF053070 Homo sapiens NADH:
gb_pr:HSY17383          -    9.00  135.74   27.63     1642  ! Y17383 Homo sapiens NDUFV1
gb_pat:AR137565         +    9.00  132.71   40.75     2517  ! AR137565 Sequence 51 from p

seq_name: gb_pr:AF304370

seq_documentation_block:
LOCUS       AF304370     2908 bp    mRNA            PRI       23-FEB-2001
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 mRNA, complete cds.
ACCESSION   AF304370
VERSION     AF304370.1  GI:10880932
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1. .2908
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     CDS             1. .2481
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24441.1"
                     /db_xref="GI:10880933"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
                     EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
```

LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT 702 a 793 c 826 g 587 t
ORIGIN

alignment_scores:
Quality: 826.00 Length: 826
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-988-687-2 x AF304370 ..

Align seg 1/1 to: AF304370 from: 1 to: 2908

```
      1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 50

     17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 100

     34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 150

     51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 200

     67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 250

     84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 300

    101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    301 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 350

    117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    351 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 400

    134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    401 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 450

    151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    451 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 500

    167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    501 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 550

    184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    551 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 600

    201 GluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluSe 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    601 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTC 650

    217 rAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGlyV 234
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    651 GAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGG 700

    234 alArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLys 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    701 TCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAG 750

    251 ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    751 AGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGG 800

    267 yThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSerI 284
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    801 GACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCA 850

    284 leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    851 TCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 900

    301 AspProGlyAlaAlaPheValValValGluCysProAspGluSerPheIl 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    901 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCAT 950

    317 eGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA 334
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    951 TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAG 1000

    334 spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal 350
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1001 ATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTG 1050

    351 AspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHi 367
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1051 GACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCA 1100

    367 sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL 384
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1101 CTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 1150

    384 ysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeu 400
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1151 AGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1200

    401 ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa 417
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1201 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGT 1250

    417 lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpG 434
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1251 TCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGC 1300

    434 lnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAla 450
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1301 AGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCG 1350

    451 LeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgSerAl 467
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1351 CTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGC 1400

    467 aGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI 484
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1401 GCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCA 1450

    484 lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1451 TCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1500

    501 AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl 517
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1501 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGG 1550

    517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1551 TGAGGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACA 1600

    534 rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp 550
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   1601 GGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGAT 1650
```

```
  551 HisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAl 567
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1651 CACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 1700

  567 aSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnL 584
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1701 ATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGC 1750

  584 euLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHis 600
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1751 TCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1800

  601 HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSe 617
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1801 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTC 1850

  617 rSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspL 634
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1851 CAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATT 1900

  634 euGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGly 650
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1901 TGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGC 1950

  651 CysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 1951 TGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC 2000

  667 rMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuI 684
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2001 CATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGA 2050

  684 leHisGluAlaThrLeuGluAspGlyLeuGluGluGluAlaValGluLys 700
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2051 TACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2100

  701 ThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAl 717
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2101 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGC 2150

  717 aGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProL 734
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2151 GGAGTTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCC 2200

  734 euPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMet 750
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2201 TCTTCAGCCCCAACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATG 2250

  751 LysValCysPheGlyAspPheProThrMetProLysLeuIleProProLe 767
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2251 AAGGTCTGCTTTGGAGACTTTCCAACAATGCCCAAGCTGATTCCCCCACT 2300

  767 uLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgArgGluL 784
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2301 GAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAGCGCAGGGAGA 2350

  784 ysArgGluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAla 800
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2351 AGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2400

  801 GlyGlyLeuGluAspGlyGluProGlnGlnLysArgAlaHisThrGluGl 817
      ||||||||||||||||||||||||||||||||||||||||||||||||||
 2401 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGA 2450

  817 uProGlnAlaLysLysValArgAlaGln 826
      |||||||||||||||||||||||||||
 2451 GCCACAGGCCAAGAAGGTCAGAGCCCAG 2478

seq_name: gb_pr:BC001939

seq_documentation_block:
LOCUS       BC001939     2997 bp    mRNA            PRI       12-JUL-2001
DEFINITION  Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:4102 IMAGE:2820640, mRNA, complete cds.
ACCESSION   BC001939
VERSION     BC001939.1  GI:12804972
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2997)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 9 Row: i Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7022621.
FEATURES             Location/Qualifiers
     source          1. .2997
                     /organism="Homo sapiens"
                     /db_xref="LocusID:60528"
                     /db_xref="taxon:9606"
                     /clone="MGC:4102 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             61. .2541
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH01939.1"
                     /db_xref="GI:12804973"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSELN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
                     EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT      740 a    806 c    853 g    598 t
ORIGIN
```

```
alignment_scores:
            Quality:  725.00              Length:    826
              Ratio:  0.879                Gaps:      0
 Percent Similarity:  99.879   Percent Identity: 99.879

alignment_block:
 US-09-988-687-2 x BC001939   ..

 Align seg 1/1  to: BC001939  from: 1  to: 2997

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      61 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 110

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     111 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 160

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     161 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 210

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     211 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 260

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     261 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 310

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     311 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 360

     101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     361 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 410

     117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     411 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 460

     134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     461 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 510

     151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     511 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 560

     167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     561 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 610

     184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     611 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 660

     201 GluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluSe 217
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     661 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTT 710

     217 rAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGlyV 234
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     711 GAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGG 760

     234 alArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLys 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     761 TCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAG 810

     251 ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl 267
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     811 AGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGG 860

     267 yThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSerI 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     861 GACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCA 910

     284 leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     911 TCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 960

     301 AspProGlyAlaAlaPheValValValGluCysProAspGluSerPheIl 317
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     961 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCAT 1010

     317 eGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA 334
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1011 TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAG 1060

     334 spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1061 ATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTG 1110

     351 AspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHi 367
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1111 GACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCA 1160

     367 sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL 384
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1161 CTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 1210

     384 ysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeu 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1211 AGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1260

     401 ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa 417
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1261 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGT 1310

     417 lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpG 434
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1311 TCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGC 1360

     434 lnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAla 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1361 AGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCG 1410

     451 LeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgSerAl 467
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1411 CTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGC 1460

     467 aGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI 484
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1461 GCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCA 1510

     484 lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1511 TCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1560

     501 AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl 517
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1561 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGG 1610

     517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1611 TGAGGGCACGTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACA 1660

     534 rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1661 GGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGAT 1710

     551 HisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAl 567
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1711 CACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 1760
```

```
     567 aSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnL 584
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1761 ATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGC 1810

     584 euLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHis 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1811 TCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1860

     601 HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSe 617
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1861 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTC 1910

     617 rSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspL 634
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1911 CAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATT 1960

     634 euGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGly 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    1961 TGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGC 2010

     651 CysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2011 TGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC 2060

     667 rMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuI 684
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2061 CATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGA 2110

     684 leHisGluAlaThrLeuGluAspGlyLeuGluGluGluAlaValGluLys 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2111 TACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2160

     701 ThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAl 717
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2161 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGC 2210

     717 aGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProL 734
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2211 GGAGTTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCC 2260

     734 euPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMet 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2261 TCTTCAGCCCCAACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATG 2310

     751 LysValCysPheGlyAspPheProThrMetProLysLeuIleProProLe 767
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2311 AAGGTCTGCTTTGGAGACTTTCCAACAATGCCCAAGCTGATTCCCCCACT 2360

     767 uLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgArgGluL 784
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2361 GAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAGCGCAGGGAGA 2410

     784 ysArgGluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAla 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2411 AGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2460

     801 GlyGlyLeuGluAspGlyGluProGlnGlnLysArgAlaHisThrGluGl 817
         ||||||||||||||||||||||||||||||||||||||||||||||||||
    2461 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGA 2510

     817 uProGlnAlaLysLysValArgAlaGln 826
         |||||||||||||||||||||||||||
    2511 GCCACAGGCCAAGAAGGTCAGAGCCCAG 2538

seq_name: gb_pr:BC004158

seq_documentation_block:
LOCUS       BC004158     3006 bp    mRNA            PRI       12-JUL-2001
DEFINITION  Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:2441 IMAGE:2820640, mRNA, complete cds.
ACCESSION   BC004158
VERSION     BC004158.1  GI:13278770
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3006)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: c Column: 13
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7022621.
FEATURES             Location/Qualifiers
     source          1. .3006
                     /organism="Homo sapiens"
                     /db_xref="LocusID:60528"
                     /db_xref="taxon:9606"
                     /clone="MGC:2441 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             70. .2550
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH04158.1"
                     /db_xref="GI:13278771"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSELN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
                     EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT      742 a    808 c    858 g    598 t
ORIGIN


alignment_scores:
            Quality:  725.00            Length:    826
              Ratio:  0.879               Gaps:      0
 Percent Similarity:  99.879   Percent Identity: 99.879

alignment_block:
 US-09-988-687-2 x BC004158   ..

 Align seg 1/1  to: BC004158  from: 1  to: 3006

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
```

```
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 70 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 119

 17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL  34
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
120 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 169

 34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly  50
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
170 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 219

 51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe  67
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
220 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 269

 67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL  84
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
270 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 319

 84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
320 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 369

101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
370 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 419

117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
420 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 469

134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
470 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 519

151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
520 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 569

167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
570 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 619

184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
620 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 669

201 GluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluSe 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
670 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTT 719

217 rAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGlyV 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
720 GAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGG 769

234 alArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLys 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
770 TCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAG 819

251 ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
820 AGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGG 869

267 yThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSerI 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
870 GACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCA 919

284 leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
920 TCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 969

301 AspProGlyAlaAlaPheValValValGluCysProAspGluSerPheIl 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
970 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCAT 1019

317 eGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1020 TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAG 1069

334 spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1070 ATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTG 1119

351 AspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHi 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1120 GACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCA 1169

367 sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1170 CTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 1219

384 ysIleGlnThrGlnLeuAsnLeuIleHisProAspIleProLeuLeu 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1220 AGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1269

401 ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa 417
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1270 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGT 1319

417 lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpG 434
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1320 TCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGC 1369

434 lnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAla 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1370 AGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCG 1419

451 LeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgSerAl 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1420 CTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGC 1469

467 aGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI 484
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1470 GCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCA 1519

484 lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1520 TCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1569

501 AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl 517
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1570 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGG 1619

517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1620 TGAGGGCACGTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACA 1669

534 rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1670 GGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGAT 1719

551 HisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAl 567
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1720 CACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 1769

567 aSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnL 584
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1770 ATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGC 1819

584 euLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHis 600
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1820 TCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1869

601 HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSe 617
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1870 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTC 1919
```

```
   617 rSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspL 634
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  1920 CAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATT 1969

   634 euGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGly 650
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  1970 TGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGC 2019

   651 CysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2020 TGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC 2069

   667 rMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuI 684
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2070 CATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGA 2119

   684 leHisGluAlaThrLeuGluAspGlyLeuGluGluGluAlaValGluLys 700
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2120 TACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2169

   701 ThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAl 717
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2170 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGC 2219

   717 aGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProL 734
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2220 GGAGTTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCC 2269

   734 euPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMet 750
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2270 TCTTCAGCCCCAACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATG 2319

   751 LysValCysPheGlyAspPheProThrMetProLysLeuIleProProLe 767
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2320 AAGGTCTGCTTTGGAGACTTTCCAACAATGCCCAAGCTGATTCCCCCACT 2369

   767 uLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgArgGluL 784
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2370 GAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAGCGCAGGGAGA 2419

   784 ysArgGluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAla 800
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2420 AGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2469

   801 GlyGlyLeuGluAspGlyGluProGlnGlnLysArgAlaHisThrGluGl 817
       ||||||||||||||||||||||||||||||||||||||||||||||||||
  2470 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGA 2519

   817 uProGlnAlaLysLysValArgAlaGln 826
       |||||||||||||||||||||||||||
  2520 GCCACAGGCCAAGAAGGTCAGAGCCCAG 2547

seq_name: gb_pr:AK001392

seq_documentation_block:
LOCUS       AK001392     2976 bp    mRNA            PRI       22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ10530 fis, clone NT2RP2000985.
ACCESSION   AK001392
VERSION     AK001392.1  GI:7022621
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2000985.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2976)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .2976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone="NT2RP2000985"
                     /clone_lib="NT2RP2"
                     /note="cloning vector: pME18SFL3; mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             75. .2639
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91666.1"
                     /db_xref="GI:7022622"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAMRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFRCKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEAL
                     QLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISPD
                     TSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERALA
                     SLGKPLHPLLVVAPNQLKAWLQQYYNQCQEVLHHISMIPAKCLQEGAEISSPAVERLI
                     SSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATLL
                     IHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFSE
                     KVGVAFDHMKVCLGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSREL
                     AGGLEDGSLSRSGPTQRSHRPRRSEPSEDLGDPELRRLCVFCPTHAPVSALLAGRS"
BASE COUNT      712 a    807 c    856 g    601 t
ORIGIN


alignment_scores:
           Quality:  430.00            Length:    753
             Ratio:  0.573               Gaps:      1
 Percent Similarity:  99.602   Percent Identity: 99.602

alignment_block:
 US-09-988-687-2 x AK001392   ..

 Align seg 1/1  to: AK001392  from: 1  to: 2976

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      75 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 124

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     125 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 174

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     175 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 224
```

```
 51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
    ||||||||||||||||||||||||||||||||||||||||||||||||||
225 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 274

 67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
    ||||||||||||||||||||||||||||||||||||||||||||||||||
275 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 324

 84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
    ||||||||||||||||||||||||||||||||||||||||||||||||||
325 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 374

101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
    ||||||||||||||||||||||||||||||||||||||||||||||||||
375 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 424

117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
    ||||||||||||||||||||||||||||||||||||||||||||||||||
425 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 474

134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
    ||||||||||||||||||||||||||||||||||||||||||||||||||
475 CAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 524

151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
    ||||||||||||||||||||||||||||||||||||||||||||||||||
525 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTATGCGGCC 574

167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
    ||||||||||||||||||||||||||||||||||||||||||||||||||
575 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 624

184 roIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSerPro 200
    ||||||||||||||||||||||||||||||||||||||||||||||||||
625 CCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 674

201 GluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluSe 217
    ||||||||||||||||||||||||||||||||||||||||||||||||||
675 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTC 724

217 rAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGlyV 234
    ||||||||||||||||||||||||||||||||||||||||||||||||||
725 GAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGG 774

234 alArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLys 250
    ||||||||||||||||||||||||||||||||||||||||||||||||||
775 TCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAG 824

251 ArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGl 267
    ||||||||||||||||||||||||||||||||||||||||||||||||||
825 AGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGG 874

267 yThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSerI 284
    ||||||||||||||||||||||||||||||||||||||||||||||||||
875 GACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCA 924

284 leThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPro 300
    ||||||||||||||||||||||||||||||||||||||||||||||||||
925 TCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 974

301 AspProGlyAlaAlaPheValValValGluCysProAspGluSerPheIl 317
    ||||||||||||||||||||||||||||||||||||||||||||||||||
975 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCAT 1024

317 eGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaA 334
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1025 TCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAG 1074

334 spAlaProValAlaLeuValValHisMetAlaProAlaSerValLeuVal 350
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1075 ATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTG 1124

351 AspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHi 367
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1125 GACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACCCAGCA 1174

367 sLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisL 384
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1175 CTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 1224

384 ysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeu 400
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1225 AGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1274

401 ThrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVa 417
    |||||||||||||||   ||||||||||||||||||||||||||||||||
1275 ACCAGTTTCCGCTGT...AAGGAGGGCCCCACCCTCAGTGTGCCCATGGT 1321

417 lGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpG 434
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1322 TCAGGGTGAATGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGC 1371

434 lnArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAla 450
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1372 AGAGGGATGCCATTATTACTTGCAATCCTGAGGAATTCATAGTTGAGGCG 1421

451 LeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgSerAl 467
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1422 CTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGGAGTGC 1471

467 aGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleI 484
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1472 GCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCA 1521

484 lePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSer 500
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1522 TCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1571

501 AlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGl 517
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1572 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGG 1621

517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1622 TGAGGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACA 1671

534 rgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAsp 550
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1672 GGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGAT 1721

551 HisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAl 567
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1722 CACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGC 1771

567 aSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnL 584
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1772 ATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGC 1821

584 euLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHis 600
    ||||||||||||||||||||||   ||||||||||||||||||||||||
1822 TCAAAGCCTGGCTCCAGCAGTACTACAACCAGTGCCAGGAGGTCCTGCAC 1871

601 HisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSe 617
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1872 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTC 1921

617 rSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspL 634
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1922 CAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATT 1971

634 euGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGly 650
    ||||||||||||||||||||||||||||||||||||||||||||||||||
1972 TGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGC 2021

651 CysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   2022 TGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACAC 2071

    667 rMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuI 684
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2072 CATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGA 2121

    684 leHisGluAlaThrLeuGluAspGlyLeuGluGluGluAlaValGluLys 700
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2122 TACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2171

    701 ThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAl 717
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2172 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGC 2221

    717 aGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProL 734
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2222 GGAGTTCATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCC 2271

    734 euPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMet 750
        ||||||||||||||||||||||||||||||||||||||||||||||||||
   2272 TCTTCAGCCCCAACTTCAGCGAGAAAGTGGGAGTTGCCTTTGACCACATG 2321

    751 LysValCys 753
        ||||||||
   2322 AAGGTCTGC 2330
```

seq_name: gb_pr:AF308698

seq_documentation_block:

```
LOCUS       AF308698     2908 bp    mRNA            PRI       27-FEB-2001
DEFINITION  Pan troglodytes ELAC2 mRNA, complete cds.
ACCESSION   AF308698
VERSION     AF308698.1  GI:10946496
KEYWORDS    .
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Pan troglodytes ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES             Location/Qualifiers
     source          1. .2908
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     CDS             1. .2481
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24920.1"
                     /db_xref="GI:10946497"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGIQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPESVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFPCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIIEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLLNILLQRERAL
                     ASLGKPFHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFN
                     EKVGVAFDHMKVCFGDFATMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT      712 a    788 c    819 g    589 t
ORIGIN
```

alignment_scores:

```
            Quality:  254.00              Length:    254
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

alignment_block:

US-09-988-687-2 x AF308698 ..

Align seg 1/1 to: AF308698 from: 1 to: 2908

```
     92 GlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIl 108
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    274 CAGAGACTCATGCAGGAGCACAAGTTAAAGGTTGCTCGCCTGGACAACAT 323

    108 ePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetI 125
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    324 ATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGCTTAAGTGGAATGA 373

    125 leLeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyPro 141
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    374 TTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGACCT 423

    142 ProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSerGlyProLe 158
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    424 CCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATT 473

    158 uLysGlyIleGluLeuAlaValArgProHisSerAlaProGluTyrGluA 175
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    474 GAAAGGAATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGG 523

    175 spGluThrMetThrValTyrGlnIleProIleHisSerGluGlnArgArg 191
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    524 ATGAAACCATGACAGTTTACCAGATCCCCATACACAGTGAACAGAGGAGG 573

    192 GlyLysHisGlnProTrpGlnSerProGluArgProLeuSerArgLeuSe 208
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    574 GGAAAGCACCAACCATGGCAGAGTCCAGAAAGGCCTCTCAGCAGGCTCAG 623

    208 rProGluArgSerSerAspSerGluSerAsnGluAsnGluProHisLeuP 225
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    624 TCCAGAGCGATCTTCAGACTCCGAGTCAAATGAAAATGAGCCACACCTTC 673

    225 roHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuValVal 241
```

```
     |||||||||||||||||||||||||||||||||||||||||||||||||
   674 CACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGACTCTTCCCTGGTCGTA 723

   242 AlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeuLy 258
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   724 GCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGCTCAA 773

   258 sAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIleI 275
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   774 AGCAAAGGAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCA 823

   275 leAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGluIle 291
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   824 TTGCTGCTGTCAAGGACGGGAAAAGCATCACTCATGAAGGAAGAGAGATT 873

   292 LeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheValVa 308
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   874 TTGGCTGAAGAGCTGTGTACTCCTCCAGATCCTGGTGCTGCTTTTGTGGT 923

   308 lValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAlaT 325
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   924 GGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTGTGAGAATGCCA 973

   325 hrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValVal 341
     |||||||||||||||||||||||||||||||||||||||||||||||||||
   974 CCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTGGTT 1023

   342 HisMetAlaPro 345
     ||||||||||||
  1024 CACATGGCCCCA 1035

seq_name: gb_pr:AF308694

seq_documentation_block:
LOCUS       AF308694     2893 bp    mRNA            PRI       27-FEB-2001
DEFINITION  Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
ACCESSION   AF308694
VERSION     AF308694.1  GI:10946488
KEYWORDS    .
SOURCE      gorilla.
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Gorilla gorilla ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES             Location/Qualifiers
     source          1. .2893
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
     gene            1. .2893
                     /gene="ELAC2"
     CDS             1. .2481
                     /gene="ELAC2"
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24916.1"
                     /db_xref="GI:10946489"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     VRLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGRHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPESVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFPCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSVQDVPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLLNILLQREQAL
                     ASLGKPLHPLLVVAPSQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFN
                     EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSGE
                     LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT      705 a    787 c    815 g    586 t
ORIGIN


alignment_scores:
            Quality:  157.00              Length:    346
              Ratio:  0.456                Gaps:      1
 Percent Similarity:  99.422   Percent Identity: 99.422

alignment_block:
 US-09-988-687-2 x AF308694   ..

 Align seg 1/1  to: AF308694  from: 1  to: 2893

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 50

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 100

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 150

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 TGCTCCGGGGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 200

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrL 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATC 250

      84 euPheAsnCysGlyGluGlyValGlnArgLeuMetGlnGluHisLysLeu 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 TCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 300
```

```
    101 LysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAs 117
        ||||||   ||||||||||||||||||||||||||||||||||||||||||
    301 AAGGTTGTTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAA 350

    117 nValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuP 134
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    351 TGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTC 400

    134 roLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAla 150
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    401 CAAAGTGTGTACTTTCTGGACCTCCACAGCTGGAAAAATACCTCGAAGCA 450

    151 IleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgPr 167
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    451 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCC 500

    167 oHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIleP 184
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    501 CCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCC 550

    184 roIleHisSerGluGlnArgArgGlyLys.HisGlnProTrpGlnSerPr 200
        ||||||||||||||||||    ||||||  ||||||||||||||||||||
    551 CCATACACAGTGAACAGAGGAG.GGGAAGGCACCAACCATGGCAGAGTCC 599

    200 oGluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGluS 217
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    600 AGAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGT 649

    217 erAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArgGly 233
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    650 CGAATGAAAATGAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGG 699

    234 ValArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeuLy 250
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    700 GTCAGGGACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAA 749

    250 sArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValG 267
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    750 GAGAGGAAACTTCTTGGTGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTG 799

    267 lyThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLysSer 283
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    800 GGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGC 849

    284 IleThrHisGluGlyArgGluIleLeuAlaGluGluLeuCysThrProPr 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    850 ATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCC 899

    300 oAspProGlyAlaAlaPheValValValGluCysProAspGluSerPheI 317
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    900 AGATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCA 949

    317 leGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAla 333
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
    950 TTCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCA 999

    334 AspAlaProValAlaLeuValValHisMetAlaPro 345
        |||||||||||||||||||||||||||||||||||
   1000 GATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCA 1035

seq_name: gb_pr:AC005277

seq_documentation_block:
LOCUS       AC005277   118788 bp    DNA             PRI       23-JUL-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.597_M_12, complete sequence.
ACCESSION   AC005277
VERSION     AC005277.1  GI:3337311
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.597_M_12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 23, 1998 this sequence version replaced gi:3335015.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Only the first 118.8 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC005274 (WICGR project
            L350).
FEATURES             Location/Qualifiers
     source          1. .118788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.597_M_12"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   795. .884
                     /rpt_family="MIR"
     repeat_region   complement(932. .1000)
                     /rpt_family="MIR"
     repeat_region   2044. .2076
                     /rpt_family="AT_rich"
     repeat_region   2237. .2713
                     /rpt_family="MER34"
     repeat_region   2718. .3477
                     /rpt_family="L2"
     repeat_region   complement(3478. .3821)
                     /rpt_family="MLT1A1"
     repeat_region   3822. .4007
                     /rpt_family="L2"
     repeat_region   complement(4015. .4281)
```

```
                     /rpt_family="MER89"
repeat_region        complement(4282. .4321)
                     /rpt_family="(TAAA)n"
repeat_region        complement(4328. .4608)
                     /rpt_family="AluSx"
repeat_region        complement(4609. .4895)
                     /rpt_family="MER89"
repeat_region        complement(4896. .5186)
                     /rpt_family="AluSp"
repeat_region        5219. .5479
                     /rpt_family="L2"
repeat_region        complement(5481. .5740)
                     /rpt_family="MER34"
repeat_region        complement(5963. .6275)
                     /rpt_family="AluSp"
repeat_region        complement(6411. .6494)
                     /rpt_family="MER5A"
repeat_region        6830. .7013
                     /rpt_family="MLT1C"
repeat_region        7014. .7318
                     /rpt_family="AluYb8"
repeat_region        7319. .7581
                     /rpt_family="MLT1C"
repeat_region        7721. .8032
                     /rpt_family="AluJb"
repeat_region        8039. .8333
                     /rpt_family="AluSq"
repeat_region        9431. .9460
                     /rpt_family="AT_rich"
repeat_region        9662. .9829
                     /rpt_family="MIR"
repeat_region        complement(11302. .11728)
                     /rpt_family="MER39"
repeat_region        complement(11824. .11884)
                     /rpt_family="LTR16C"
repeat_region        complement(11928. .12108)
                     /rpt_family="LTR16C"
repeat_region        12453. .12725
                     /rpt_family="AluJb"
repeat_region        13377. .13785
                     /rpt_family="MSTB"
repeat_region        14118. .14153
                     /rpt_family="(CA)n"
repeat_region        complement(15320. .15620)
                     /rpt_family="AluJo"
repeat_region        15683. .15759
                     /rpt_family="(CA)n"
repeat_region        complement(15838. .15893)
                     /rpt_family="MER91A"
repeat_region        complement(15915. .16080)
                     /rpt_family="MER91A"
repeat_region        complement(16474. .16544)
                     /rpt_family="L2"
repeat_region        17907. .17920
                     /rpt_family="L1ME"
repeat_region        18271. .18311
                     /rpt_family="L2"
repeat_region        complement(18368. .18529)
                     /rpt_family="AluJb"
repeat_region        complement(18530. .18560)
                     /rpt_family="(CA)n"
repeat_region        complement(18561. .18681)
                     /rpt_family="AluJb"
repeat_region        18760. .19065
                     /rpt_family="AluSq"
repeat_region        complement(19638. .19896)
                     /rpt_family="MIR"
repeat_region        complement(19971. .20260)
                     /rpt_family="L1ME3A"
repeat_region        20263. .20552
                     /rpt_family="MER58B"
repeat_region        complement(20553. .20816)
                     /rpt_family="L1ME3A"
repeat_region        complement(20821. .21069)
                     /rpt_family="L1M4"
repeat_region        complement(21071. .21100)
                     /rpt_family="(TAAA)n"
repeat_region        complement(21101. .21338)
                     /rpt_family="AluSx"
repeat_region        complement(21348. .21598)
                     /rpt_family="L1M4"
repeat_region        21799. .22104
                     /rpt_family="AluJo"
repeat_region        22151. .22562
                     /rpt_family="MSTA"
repeat_region        23978. .24287
                     /rpt_family="AluSq"
repeat_region        24852. .25199
                     /rpt_family="L2"
repeat_region        complement(25697. .25985)
                     /rpt_family="MLT1D"
unsure               25816. .25895
                     /note="Single-stranded coverage."
repeat_region        complement(25994. .26099)
                     /rpt_family="purine-rich"
repeat_region        complement(26122. .26291)
                     /rpt_family="MLT1D"
repeat_region        26312. .26337
                     /rpt_family="(CAAAA)n"
repeat_region        26456. .26480
                     /rpt_family="(GGGAA)n"
repeat_region        complement(26576. .26781)
                     /rpt_family="MER6B"
repeat_region        complement(26912. .26990)
                     /rpt_family="MER21B"
repeat_region        26991. .27290
                     /rpt_family="AluSx"
repeat_region        complement(27291. .27975)
                     /rpt_family="MER21B"
repeat_region        28034. .28197
                     /rpt_family="LTR28"
repeat_region        28266. .28795
                     /rpt_family="LTR28"
repeat_region        28843. .28916
                     /rpt_family="LTR28"
repeat_region        28920. .29005
                     /rpt_family="MIR"
repeat_region        29140. .29233

alignment_scores:
            Quality:   82.00          Length:     82
              Ratio:  1.000             Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AC005277   ..

 Align seg 1/1  to: AC005277  from: 1  to: 118788

       1 MetTrpAlaLeuCysSerLeuLeuArgSerAlaAlaGlyArgThrMetSe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   78562 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTC 78611

      17 rGlnGlyArgThrIleSerGlnAlaProAlaArgArgGluArgProArgL 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   78612 GCAGGGACGCACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCA 78661

      34 ysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerGly 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   78662 AGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 78711

      51 CysSerGlyGlyProAsnThrValTyrLeuGlnValValAlaAlaGlySe 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   78712 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAG 78761

      67 rArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnArg 82
```

```
          ||||||||||||||||||||||||||||||||||||||||||||
   78762 CCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGG 78807

seq_name: gb_ro:AF308696

seq_documentation_block:
LOCUS       AF308696     2712 bp    mRNA            ROD       27-FEB-2001
DEFINITION  Mus musculus ELAC2 mRNA, complete cds.
ACCESSION   AF308696
VERSION     AF308696.2  GI:11992378
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Mouse ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
REFERENCE   4  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-DEC-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
  REMARK    Sequence update by submitter
COMMENT     On Dec 26, 2000 this sequence version replaced gi:10946492.
FEATURES             Location/Qualifiers
     source          1. .2712
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /dev_stage="fetus"
     CDS             15. .2510
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24918.2"
                     /db_xref="GI:11992379"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPSKDPLRHLRTREKRGPGP
                     GGPNTVYLQVVAAGGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERRCGKQQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     FPQLTSFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTPDCNTDEFIAEALELP
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     LETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAGFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPLKALFAGDIEEMVERREKRELRLVRAALLTQQADSP
                     EDREPQQKRAHTDEPHSPQSKKESVANTLGARV"
BASE COUNT      659 a    741 c    748 g    564 t
ORIGIN


alignment_scores:
           Quality:   51.00            Length:     51
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AF308696   ..

 Align seg 1/1  to: AF308696  from: 1  to: 2712

     123 GlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValLeuSe 139
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     357 GGAATGATTTTAACTTTAAAGGAAACCGGGCTTCCCAAATGTGTTCTGTC 406

     139 rGlyProProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSerG 156
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     407 TGGACCACCACAGCTGGAGAAATATCTAGAAGCAATCAAAATATTTTCTG 456

     156 lyProLeuLysGlyIleGluLeuAlaValArgProHisSerAlaProGlu 172
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     457 GTCCATTGAAAGGAATAGAACTGGCCGTGCGGCCTCACTCTGCACCAGAA 506

     173 Tyr 173
         |||
     507 TAC 509

seq_name: gb_pr:AF304371S2

seq_documentation_block:
LOCUS       AF304371S2    740 bp    DNA             PRI       19-OCT-2000
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 gene, exon 17 and partial cds.
ACCESSION   AF304369
VERSION     AF304369.1  GI:10880929
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
```

```
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     exon            301. .439
                     /number=17
     mRNA            <301. .>439
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
     CDS             <301. .>439
                     /codon_start=2
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24440.1"
                     /db_xref="GI:10880931"
                     /translation="PDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADH
                     HT"
BASE COUNT      154 a    188 c    173 g    225 t
ORIGIN


alignment_scores:
            Quality:   47.00            Length:     47
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AF304371S2   ..

 Align seg 1/1  to: AF304371S2  from: 1  to: 740

     507 SerProAspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGl 523
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     299 AGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTGGGCA 348

     523 nLeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuA 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     349 GCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGG 398

     540 laAlaValPheValSerHisLeuHisAlaAspHisHisThr 553
         |||||||||||||||||||||||||||||||||||||||
     399 CTGCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACG 439

seq_name: gb_pr:AF304371S1

seq_documentation_block:
LOCUS       AF304371S1    721 bp    DNA             PRI       19-OCT-2000
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 gene, exon 7 and partial cds.
ACCESSION   AF304371
VERSION     AF304371.1  GI:10880928
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 721)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1. .721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     exon            301. .420
                     /number=7
     mRNA            <301. .>420
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
     CDS             <301. .>420
                     /codon_start=3
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24439.1"
                     /db_xref="GI:10880930"
                     /translation="EQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPH"
BASE COUNT      201 a    143 c    184 g    193 t
ORIGIN


alignment_scores:
            Quality:   40.00            Length:     40
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AF304371S1   ..

 Align seg 1/1  to: AF304371S1  from: 1  to: 721

     188 GluGlnArgArgGlyLysHisGlnProTrpGlnSerProGluArgProLe 204
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     303 GAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCAGAAAGGCCTCT 352

     204 uSerArgLeuSerProGluArgSerSerAspSerGluSerAsnGluAsnG 221
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     353 CAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTCGAATGAAAATG 402

     221 luProHisLeuProHisGly 227
         ||||||||||||||||||||
     403 AGCCACACCTTCCACATGGT 422
```

```
seq_name: gb_ro:AF348157

seq_documentation_block:
LOCUS       AF348157    34593 bp    DNA             ROD       04-APR-2001
DEFINITION  Mus musculus putative prostate cancer susceptibility protein
            (Elac2) gene, complete cds, alternatively spliced.
ACCESSION   AF348157
VERSION     AF348157.1  GI:13540341
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Mouse Elac2-containing genomic DNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 34593)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2001) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1. .34593
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     mRNA            join(<8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .>30639)
                     /gene="Elac2"
                     /note="alternatively spliced 24 exon form"
                     /product="putative prostate cancer susceptibility protein"
     mRNA            join(<8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .30624,
                     31242. .>31277)
                     /gene="Elac2"
                     /note="alternatively spliced 25 exon form"
                     /product="putative prostate cancer susceptibility protein"
     gene            <8470. .>31277
                     /gene="Elac2"
     CDS             join(8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .30639)
                     /gene="Elac2"
                     /note="24 exon form"
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAK29421.1"
                     /db_xref="GI:13540343"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTREKRGPGP
                     GGPNTVYLQVVAAGGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERRCGKQQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     FPQLTSFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTLDCNTDEFIAEALELP
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     LETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPLKALFAGDIEEMVERREKRELRLVRAALLTQQADSP
                     EDREPQQKRAHTDEPHSPQSKKVRAQ"
     CDS             join(8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .30624,
                     31242. .31277)
                     /gene="Elac2"
                     /note="25 exon form"
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAK29420.1"
                     /db_xref="GI:13540342"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTREKRGPGP
                     GGPNTVYLQVVAAGGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERRCGKQQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     FPQLTSFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTLDCNTDEFIAEALELP
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     LETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPLKALFAGDIEEMVERREKRELRLVRAALLTQQADSP
                     EDREPQQKRAHTDEPHSPQSKKESVANTLGARV"
BASE COUNT     8641 a    8030 c    8142 g    9780 t
ORIGIN


alignment_scores:
           Quality:   26.00            Length:     26
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AF348157   ..

 Align seg 1/1  to: AF348157  from: 1  to: 34593

     677 GlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGlyLe 693
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   29100 GGGAAAGATGCCACCCTCCTGATACATGAAGCCACTCTGGAGGATGGCTT 29149

     693 uGluGluGluAlaValGluLysThrHis 702
         |||||||||||||||||||||||||||
   29150 GGAAGAGGAAGCAGTAGAGAAGACACAC 29177

seq_name: gb_pl:SPAC1D4

seq_documentation_block:
LOCUS       SPAC1D4     31201 bp    DNA             PLN       15-OCT-1999
DEFINITION  S.pombe chromosome I cosmid c1D4.
```

```
ACCESSION   Z69239
VERSION     Z69239.1  GI:1177333
KEYWORDS    byr1; cct2; CDP-alcohol phosphatidyltransferase; chaperonin;
            conjugation and sporulation; csk1; cyclin suppressing protein
            kinase; DNA repair helicase; golgi peripheral membrane protein;
            protein kinase; rad15; rhp3; serine threonine protein kinase; ste1;
            TCP1 beta subunit homologue; tf2 LTR.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (bases 1 to 31201)
  AUTHORS   Lye,G., Churcher,C.M., Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-1995) Schizosaccharomyces pombe chromosome I
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT     Notes:
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_pombe)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
            pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid c1D4 is overlapped at the 5' end by cosmid c1F3 and at the
            3' end by cosmid c22F3.
FEATURES             Location/Qualifiers
     source          1. .31201
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /clone="cosmid c1D4"
                     /map="IL"
     misc_feature    1. .133
                     /note="the sequence of this region is derived from cosmid
                     c 1F3, the true end of c1D4 is at position 134, while the
                     true end of c1F3 is at position 4137 in this sequence"
     gene            7. .864
                     /gene="SPAC1F3.11"
                     /note="SPAC1D4.01"
     CDS             7. .864
                     /gene="SPAC1F3.11"
                     /note="SPAC1D4.01, len:285"
                     /codon_start=1
                     /product="hypothetical 33.0 kd protein"
                     /protein_id="CAB56125.1"
                     /db_xref="GI:5912524"
                     /translation="MRQIFGTCTGYMNGIYFQEGLLEIPTTKSIRMNSIHIKKKSNRS
                     FRRRKVFGNEKEFDLEELDDNDIRLRQALEATKRRKIRNSIIGINAEKLLNQETKKEK
                     QLNTANEPHEANDQTSAQSSKLIEAQLPTVEDRFAKQTNEVDINTHLLNFVEKKLKQE
                     RLAQNYSENGETNALNTKNESTVQNIKNSLHPNEHSFIRDAAALGAIREVDLGIISTD
                     VDNLKNGRKRQKKRARMKEKLDSKALRTSEDAARDEFIEKMLKPISQDEESKGIYRRF
                     RVYKDGTQD"
     gene            complement(join(1088. .1636,1686. .1873,1970. .2090,
                     2137. .2316))
                     /gene="SPAC1D4.02c"
     CDS             complement(join(1088. .1636,1686. .1873,1970. .2090,
                     2137. .2316))
                     /gene="SPAC1D4.02c"
                     /note="SPAC1D4.02c, len:345, SIMILARITY:Rattus norvegicus,
                     O35254, golgi peripheral membrane protein p65., (451 aa),
                     fasta scores: opt: 493, E():4.7e-20, (34.4% identity in
                     305 aa)"
                     /codon_start=1
                     /label=SPAC1D4.02c
                     /product="putative golgi peripheral membrane protein"
                     /protein_id="CAB56126.1"
                     /db_xref="GI:5912525"
                     /translation="MFGGLKNFIKEKSEALAGIHRESDESCGFRVLKVENDSKAYNAR
                     IESYYDFITAVNGILLNGDPSMFMALLRDSSPEVTLEVFSLKGQITRKVNIKINSDEK
                     IGMVLQWASIAPAVDAIWHILNVIDDSPVARASLVPYEDYIVGTPEGMMTGEKALSDL
                     IESHLNRPLRLYIYNHYRDSTRQVTIVPNRHWGGNGAIGCGVGHGVLHRLPAPLSGPP
                     PQPGDIVFSNPMLGGPDHKVSQPSETENFLPTPEPPKIASANAGSSNEISIPHYQRHK
                     KSHKGAIQDSSIQSYLDEEEKLSRELDHKTKDASSTNDSQTTPLPPPPPVAVNSTNDE
                     SAPQNEELVKN"
misc_feature         complement(1637. .1651)
                     /note="splice branch and acceptor sequence,
                     ctaacaatttaacag"
misc_feature         complement(1680. .1685)
                     /note="splice donor sequence, gtatgt"
misc_feature         complement(1874. .1886)
                     /note="splice branch and acceptor sequence, ctaacgactacag"
misc_feature         complement(1964. .1969)
                     /note="splice donor sequence, gtaagt"
misc_feature         complement(2091. .2108)
                     /note="ctaaagtgtttttcacag, splice branch and acceptor"
misc_feature         complement(2131. .2136)
                     /note="gtatga, splice donor sequence"
gene                 complement(join(2574. .2820,2907. .4201))
                     /gene="SPAC1D4.03c"
CDS                  complement(join(2574. .2820,2907. .4201))
                     /gene="SPAC1D4.03c"
                     /note="SPAC1D4.03c, len:513, SIMILARITY:Fugu rubripes,
                     Q9YGN1, sand protein., (520 aa), fasta scores: opt: 609,
                     E():7.2e-32, (28.5% identity in 407 aa)"
                     /codon_start=1
                     /label=SPAC1D4.03c
                     /product="hypothetical protein"
                     /protein_id="CAA93212.1"
                     /db_xref="GI:1177336"
                     /db_xref="SWISS-PROT:Q10150"
                     /translation="MEPTSEHSSIKEEVENDNVHRSHESECGSLLLNPGNVLMAAPSV
                     SEDDQEVSRSTPELRSHVENVEQLLSDILHDNSSPLNVSTSVSSSSNNTAVDEIIKLL
                     SLISFDLAKQKRTYLIFSSSGKPVFSNIVDDSIEPSTVGALQAIISSFEVSKEELTSF
                     STFSNVIVVLSKNPLYLVGVSPSTTLSAAYLLSELNLLYCQILTGVTAKAMQLTLNSR
                     PNFDLRRLIGSNEQFLKELCDQLNDYELVPTLNAISPLPLRSSFRDQLSQLLLRETPK
                     SLLFTFIAIRGRLVCMVKAKKLLLHANDLYLLFLSLFRTQSFNDSMEHWVPVCFPTLN
                     PDAYIYIYSYFLCKDTVLIMGSSESGVFFEMQSVKCKVAQEIQDHGWLKKLIYCEEMD
                     RTTPRNPGSPCISHYLFYSKKYSQFYTPGYSFSTPNFNTRTLYAIYASLHDQAFHKKN
                     SFSINMTVHESLLLFTWSTASFDFHCIANATTSSQLLIANVNKILRWIRREENRLFIQ
                     TNLSF"
misc_feature         complement(2821. .2836)
                     /note="splice branch and acceptor sequence,
                     ctaactatgttttta g"
misc_feature         complement(2901. .2906)
                     /note="splice donor sequence, gtacgt"
gene                 join(5400. .5402,5439. .5570,5618. .7066)
                     /gene="SPAC1D4.04"
                     /note="cct2"
CDS                  join(5400. .5402,5439. .5570,5618. .7066)
                     /gene="SPAC1D4.04"
                     /note="SPAC1D4.04, len:527"
                     /codon_start=1
                     /product="probable t-complex protein 1, beta subunit"
                     /protein_id="CAA93213.1"
                     /db_xref="GI:1177337"
                     /db_xref="SWISS-PROT:Q10147"
                     /translation="MSLNPHQIFNESGIQERGENARLSSFVGAIAVGDLVKSTLGPKG
                     MDKILQSNSSGDIVVTNDGATILKSIALDNAAAKVLVNISKVQDDEVGDGTTSVCVFA
                     AELLRQAEIMVNAKIHPQVIIDGYRIATKTAIDALRASSIDNSSDPAKFRSDLENIAR
```

```
                     TTLSSKILSQNKNHFAQLAVDAVLRLKGSTNLDNIQIIKILGGKLDDSFLDEGFILNK
                     TIGVNCPKVMENANILIANTAMDTDKVKVFGARVRVDTTGKLAELERAEREKMKAKVE
                     KIKSHNINCFINRQLIYNWPEQLFADAGIMSIEHADFDGIERLSLVTGGEIASTFDHP
                     ELVKLGHCKKIEEIIIGEDKMIKFSGVEAGEACTIVLRGATHQLLDESERAIHDALAV
                     LSQTVAESRVTLGGGCAEMLMAKAVEEAATHEPGKKAVAVSAFAKALSQLPTILADNA
                     GFDSSELVAQLKAAHYDGNDTMGLDMDEGEIADMRAKGILEALKLKQAVVSSGSEGAQ
                     LLLRVDTILKAAPRPRERM"
     misc_feature    5425. .5438
                     /note="splice branch and acceptor sequence,
                     tactaacgatatag"
     misc_feature    join(5529. .5570,5618. .7045)
                     /gene="SPAC1D4.04"
                     /note="Match to PF00118 cpn60_TCP1, TCP-1/cpn60 chaperonin
                     family Score 590.77"
     misc_feature    5571. .5576
                     /note="splice donor sequence, gtaagg"
     misc_feature    5604. .5617
                     /note="splice branch and acceptor sequence,
                     ttaacgatttatag"
     misc_feature    5741. .5767
                     /gene="SPAC1D4.04"
                     /note="PS00995 Chaperonins TCP-1 signature 3"
     gene            complement(join(7282. .8366,8419. .8497))
                     /gene="SPAC1D4.05c"
     CDS             complement(join(7282. .8366,8419. .8497))
                     /gene="SPAC1D4.05c"
                     /note="SPAC1D4.05c, len:387, LOW SIMILARITY:Mus musculus,
                     Q9Z0U0, polytropic murine leukamia virus receptor sygl.,
                     (695 aa), fasta scores: opt: 261, E():1.5e-10, (25.1%
                     identity in 398 aa) also shows low similarity to,
                     Saccharomyces cerevisiae, ERD1_YEAST, ERD1 protein., (362
                     aa), fasta scores: opt: 187, E(): 1e-05, (23.6% identity
                     in 382 aa)
                     SPAC1D4.05c, len:387, SIMILARITY:Mus musculus, Q9Z0U0,
                     polytropic murine leukamia virus receptor sygl., (695 aa),
                     fasta scores: opt: 261, E():1.5e-10, (25.1% identity in
                     398 aa)"
                     /codon_start=1
                     /label=SPAC1D4.05c
                     /product="hypothetical protein"
                     /protein_id="CAA93214.1"
                     /db_xref="GI:1177338"
                     /db_xref="SWISS-PROT:Q10151"
                     /translation="MDLEVVEFPLHHKLALPFRIGLLVIVGTWLWSVCYHLIYILNRY
                     QPISPNPRGSLNSKWYHLLQIPLSNRHTDLEENTEFKANLVSPVDFHAGYCFAAILSI
                     SWATGFILFLKKTQGDIGGLYSHPIYPLLWVITAFILIVFPFFPWRYRSSQRGLRKSII
                     RVFLFFQADFRSPYKDFIVSEIFTSYAKALGDFYIFGCVLQGHISKFTLRPDLKCDGT
                     FFVPLAMAYPFIVAILQCLHYGLSRRKHTFKINLLSALKHATALPVIYLSAIIHAKQT
                     KFTLTSGHGYLFWLWILSALLSSAYTFLWDVFIDWRIRFPFHKSINHKRFPMFIYAIG
                     CFINFILRVTWSMKLHPRLHQFHEYEMGIFSFEMLEILRRFLWLFFHLDAISS"
     misc_feature    complement(8367. .8389)
                     /note="splice branch and acceptor sequence,

alignment_scores:
            Quality:   11.00            Length:     11
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x SPAC1D4   ..

 Align seg 1/1  to: SPAC1D4  from: 1  to: 31201

     681 ThrLeuLeuIleHisGluAlaThrLeuGluAsp 691
         |||||||||||||||||||||||||||||||||
   16609 ACTTTATTAATTCACGAAGCTACTTTGGAAGAC 16641

seq_name: gb_htg:AC036490

seq_documentation_block:
LOCUS       AC036490      881 bp    DNA             HTG       07-APR-2000
DEFINITION  Giardia intestinalis clone EJ0814 strain WB-C6, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC036490
VERSION     AC036490.1  GI:7524276
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Giardia intestinalis.
  ORGANISM  Giardia intestinalis
            Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
REFERENCE   1  (bases 1 to 881)
  AUTHORS   Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
  TITLE     Giardia: a model for ancient eukaryotic genome analysis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 881)
  AUTHORS   Eakin,N.Q., Morrison,H.G., McArthur,A.G., Nixon,J., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Josephine Bay Paul Center for Comparative
            Molecular Biology and Evolution, Marine Biological Laboratory, 7
            MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT     * NOTE: This record contains 1 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *         1      881: contig of 881 bp in length.
FEATURES             Location/Qualifiers
     source          1. .881
                     /organism="Giardia intestinalis"
                     /strain="WB-C6"
                     /db_xref="taxon:5741"
                     /clone="EJ0814"
BASE COUNT      217 a    238 c    189 g    237 t
ORIGIN


alignment_scores:
            Quality:   10.00            Length:     10
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AC036490   ..

 Align seg 1/1  to: AC036490  from: 1  to: 881

     223 HisLeuProHisGlyValSerGlnArgArg 232
         ||||||||||||||||||||||||||||||
     433 CATTTACCTCATGGAGTCTCTCAGAGAAGG 462

seq_name: gb_htg:AC031419

seq_documentation_block:
LOCUS       AC031419      938 bp    DNA             HTG       02-APR-2000
DEFINITION  Giardia intestinalis clone EI1687 strain WB-C6, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC031419
VERSION     AC031419.1  GI:7386746
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Giardia intestinalis.
  ORGANISM  Giardia intestinalis
            Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
REFERENCE   1  (bases 1 to 938)
  AUTHORS   Morrison,H.G., McArthur,A.G., Nixon,J., Eakin,N.Q., Kim,U.,
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
  TITLE     Giardia: a model for ancient eukaryotic genome analysis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 938)
  AUTHORS   Eakin,N.Q., Morrison,H.G., McArthur,A.G., Nixon,J., Kim,U.,
```

```
            Crocker,M.C., Hinkle,G., Holder,M.E. and Sogin,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-APR-2000) Josephine Bay Paul Center for Comparative
            Molecular Biology and Evolution, Marine Biological Laboratory, 7
            MBL Street, Woods Hole, MA 02543-1015, USA
COMMENT     * NOTE: This record contains 1 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *         1      938: contig of 938 bp in length.
FEATURES              Location/Qualifiers
     source           1. .938
                      /organism="Giardia intestinalis"
                      /strain="WB-C6"
                      /db_xref="taxon:5741"
                      /clone="EI1687"
BASE COUNT      246 a    244 c    198 g    250 t
ORIGIN


alignment_scores:
            Quality:   10.00              Length:     10
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AC031419   ..

 Align seg 1/1  to: AC031419  from: 1  to: 938

     223 HisLeuProHisGlyValSerGlnArgArg 232
         |||||||||||||||||||||||||||||||
     363 CATTTACCTCATGGAGTCTCTCAGAGAAGG 392

seq_name: gb_pl':AC022354

seq_documentation_block:
LOCUS       AC022354    46335 bp    DNA             PLN       25-OCT-2000
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F9I5 genomic sequence,
            complete sequence.
ACCESSION   AC022354
VERSION     AC022354.1  GI:6850338
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 46335)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M.-I., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F9I5 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46335)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 46335)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     Address all correspondence to:at@tigr.org
            BAC clone F9I5 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES              Location/Qualifiers
     source           1. .46335
                      /organism="Arabidopsis thaliana"
                      /cultivar="Columbia"
                      /db_xref="taxon:3702"
                      /chromosome="1"
                      /clone="F9I5"
     mRNA             complement(<126. .>334)
                      /gene="F9I5.18"
     gene             complement(126. .334)
                      /gene="F9I5.18"
                      /note="similar to HD-zip (Arabidopsis thaliana)
                      GI:1149569"
     CDS              complement(<128. .334)
                      /gene="F9I5.18"
                      /codon_start=1
                      /product="HD-zip-like protein, 3' partial"
                      /protein_id="AAG26011.1"
                      /db_xref="GI:10998865"
                      /translation="MAMSCKDGKLGCLDNGKYVRYTPEQVEALERLYHDCPKPSSIRR
                      QQLIRECPILSNIEPKQIKVWFQNR"
     repeat_region    1002. .1048
                      /rpt_family="(GAGAA)n"
     repeat_region    complement(1169. .1195)
                      /rpt_family="AT_rich"
     repeat_region    complement(1562. .1608)
                      /rpt_family="AT_rich"
     repeat_region    complement(2125. .2154)
                      /rpt_family="(GAAA)n"
     repeat_region    complement(3622. .3728)
                      /rpt_family="AT_rich"
     repeat_region    6914. .6989
                      /rpt_family="(TTAAA)n"
     mRNA             complement(join(<7284. .7610,7681. .8652,8869. .9016,
                      9063. .9373,9467. .9826,9904. .9947,10041. .10095,
                      10177. .10221,10306. .10395,10478. .>10639))
                      /gene="F9I5.1"
     gene             complement(7284. .10639)
                      /gene="F9I5.1"
     CDS              complement(join(7284. .7610,7681. .8652,8869. .9016,
                      9063. .9373,9467. .9826,9904. .9947,10041. .10095,
                      10177. .10221,10306. .10395,10478. .10639))
                      /gene="F9I5.1"
                      /codon_start=1
                      /product="unknown protein"
                      /protein_id="AAF29402.1"
                      /db_xref="GI:6850339"
                      /translation="MENNEATNGSKSSSNSFVFNKRRAEGFDITDKKKRNLERKSQKL
                      NPTNTIAYAQILGTGMDTQDTSSSVLLFFDKQRFIFNAGEGLQRFCTEHKIKLSKIDH
                      VFLSRVCSETAGGLPGLLLTLAGIGEEGLSVNVWGPSDLNYLVDAMKSFIPRAAMVHT
```

```
                RSFGPSSTPDPIVLVNDEVVKISAIILKPCHSEEDSGNKSGDLSVVYVCELPEILGKF
                DLEKAKKVFGVKPGPKYSRLQSGESVKSDERDITVHPSDVMGPSLPGPIVLLVDCPTE
                SHAAELFSLKSLESYYSSPDEQTIGAKFVNCIIHLSPSSVTSSPTYQSWMKKFHLTQH
                ILAGHQRFLPLLIIVSHQKTVRKNMAFPILKASSRIAARLNYLCPQFFPAPGFWPSQL
                TDNSIIDPTPSNKFNLRPVAIRGIDRSCIPAPLTSSEVVDELLSEIPEIKDKSEEIKQ
                FWNKQHNKTIIEKLWLSECNTVLPNCLEKIRRDDMEIVILGTGSSQPSKYRNVSAIFI
                DLFSRGSLLLDCGEGTLGQLKRRYGLDGADEAVRKLRCIWISHIHADHHTGLARILAL
                RSKLLKGVTHEPVIVVGPRPLKRFLDAYQRLEDLDMEFLDCRSTTATSWASLESGGEA
                EGSLFTQGSPMQSVFKRSDISMDNSSVLLCLKNLKKVLSEIGLNDLISFPVVHCPQAY
                GVVIKAAERVNSVGEQILGWKMVYSGDSRPCPETVEASRDATILIHEATFEDALIEEA
                LAKNHSTTKEAIDVGSAANVYRIVLTHFSQRYPKIPVIDESHMHNTCIAFDLMSINMA
                DLHVLPKVLPYFKTLFRDEMVEDEDADDVAMDDLKEEAL"
mRNA            complement(join(<11317. .11583,12194. .>12301))
                /gene="F9I5.3"
gene            complement(11317. .12301)
                /gene="F9I5.3"
                /note="similar to delta tonoplast integral protein
                GI:1145697 (Arabidopsis thaliana)"
CDS             complement(join(11317. .11583,12194. .12301))
                /gene="F9I5.3"
                /codon_start=1
                /product="aquaporin, putative"
                /protein_id="AAF29403.1"
                /db_xref="GI:6850340"
                /translation="MAVDEVSIFIGVASLTVSVWWFINKFTYLEAVPTSNAIPIHSVA
                VRVGSTQRVVMEIIITFALVYTVYATAIDSNNGTLGTIAPLAIRLIVGANILAAGPFS
                GGPMNPGRSFGSSLAVGNFSGH"
repeat_region   13176. .13409
                /rpt_family="novel repeat from T25M19 22823 to 23610 788
                nt"
repeat_region   13177. .13298
                /rpt_family="novel repeat from T25M19 22823 to 23610 788
                nt"
repeat_region   complement(14874. .14953)
                /rpt_family="AT_rich"
repeat_region   complement(16191. .16271)
                /rpt_family="AT_rich"
repeat_region   complement(17443. .17472)
                /rpt_family="AT_rich"
repeat_region   complement(18141. .18182)
                /rpt_family="AT_rich"
repeat_region   complement(18431. .18451)
                /rpt_family="AT_rich"
repeat_region   complement(18883. .18912)
                /rpt_family="(GAAA)n"
repeat_region   19668. .19760
                /rpt_family="(A)n"
repeat_region   complement(20768. .20803)
                /rpt_family="AT_rich"
repeat_region   complement(20931. .20985)
                /rpt_family="(GAA)n"
mRNA            join(<21044. .21140,21233. .21450,23180. .23739,24098. .25251)
                /gene="F9I5.4"
gene            21044. .25251
                /gene="F9I5.4"
                /note="similar to peptide transporter PTR2-B SP:P46032
                (Arabidopsis thaliana (Mouse-ear cress))"
CDS             join(21044. .21140,21233. .21450,23180. .23739,24098. .25046)
                /gene="F9I5.4"
                /codon_start=1
                /product="peptide transporter, putative"
                /protein_id="AAF29404.1"
                /db_xref="GI:6850341"
                /translation="MENPPNETEAKQIQTNEGKKTKGGIITMPFIIANEAFEKVASYG
                LLPNMIMYLIRDYRFGVAKGTNVLFMWSAASNFTPLLGAFLSDSYLGRFLTISIASLS
                SFLGMVLLWLTAMLPQVKPSPCDPTAAGSHCGSSTASQLALLYSAFALISIGSGGIRP
                CSLAFGADQLDNKENPKNERVLESFFGWYYASSAVAVLIAFTGIVYIQEHLGWKIGFG
                VPAVLMLIAALLFILASPLYVTRGVTKSLFTGLAQAIVAAYKKRKLSLPDHHDSFDCY
                YHMKDSEIKAPSQKLRFLNKACLISNREEEIGSDGFALNPWRLCTTDKVEELKALIKV
                IPIWSTGIMMSINTSQSSFQLLQATSMDRRLSRHGSSFQVPAGSFGMFTIIALALWVI
                LYDRAVIPLASKIRGRPFRLSVKLRMGLGLFMSFLAMAISAMVESFRRKKAISQGYAN
                NSNAVVDISAMWLVPQYVLHGLAEALTAIGQTEFFYTEFPKSMSSIAASLFGLGMAVA
                SLLASVVLNAVNELTSRNGKESWVSDNINKGHYNYYYWVLAIMSFINVIYYVICSWSY
                GPLVDQVRNGRVNGVREEEELIDIVGKGFEKEDLSPVVKTN"
repeat_region   complement(21187. .21225)
                /rpt_family="AT_rich"
repeat_region   complement(21969. .21990)
                /rpt_family="AT_rich"
repeat_region   complement(26156. .26179)
                /rpt_family="(TA)n"
repeat_region   complement(26608. .26641)
                /rpt_family="AT_rich"
repeat_region   complement(27450. .27491)
                /rpt_family="(TA)n"
mRNA            complement(join(<28479. .28732,29099. .29311,29402. .29473,
                29582. .>29780))
                /gene="F9I5.19"
gene            complement(28479. .29780)
                /gene="F9I5.19"
CDS             complement(join(28644. .28732,29099. .29311,29402. .29473,
                29582. .29780))
                /gene="F9I5.19"
                /note="tRNA-Asn"
                /codon_start=1
                /product="unknown protein"
                /protein_id="AAF29406.2"
                /db_xref="GI:10998866"
                /translation="MGRVTTPSEEDSNNGLPVQQPGTPNQRTRVPVSQFAPPNYQQAN

alignment_scores:
            Quality:   10.00              Length:     10
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-988-687-2 x AC022354/rev   ..

 Align seg 1/1  to reverse of: AC022354  from: 1  to: 46335

     511 SerLeuLeuLeuAspCysGlyGluGlyThr 520
         |||||||||||||||||||||||||||||||
    8325 AGTCTTCTCCTAGATTGTGGTGAAGGAACC 8296
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 29, 2002, 10:20:53 ; Search time 14.18 Seconds
(without alignments)
1310.841 Million cell updates/sec

Title: US-09-988-687-2
Perfect score: 826
Sequence: 1 MWALCSLLRSAAGRTMSQGR.....EPQQKRAHTEEPQAKKVRAQ 826

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9	1.1	838	4	US-09-315-794-52	Sequence 52, Appl
2	9	1.1	838	4	US-09-389-341-52	Sequence 52, Appl
3	8	1.0	138	3	US-08-930-894-5	Sequence 5, Appli
4	7	0.8	40	4	US-08-981-392-33	Sequence 33, Appl
5	7	0.8	250	2	US-08-685-992-7	Sequence 7, Appli
6	7	0.8	250	2	US-09-144-925-7	Sequence 7, Appli
7	7	0.8	289	1	US-08-036-210-13	Sequence 13, Appl
8	7	0.8	289	2	US-08-449-609-13	Sequence 13, Appl
9	7	0.8	328	2	US-08-878-989-7	Sequence 7, Appli
10	7	0.8	328	4	US-09-272-796-7	Sequence 7, Appli
11	7	0.8	432	2	US-08-677-049-8	Sequence 8, Appli
12	7	0.8	460	2	US-08-677-049-10	Sequence 10, Appl
13	7	0.8	559	1	US-08-320-559-31	Sequence 31, Appl
14	7	0.8	559	3	US-08-545-860D-31	Sequence 31, Appl
15	7	0.8	559	5	PCT-US94-04496-31	Sequence 31, Appl
16	7	0.8	872	1	US-08-451-715A-8	Sequence 8, Appli
17	7	0.8	928	2	US-08-841-483-4	Sequence 4, Appli
18	7	0.8	928	4	US-09-382-911-4	Sequence 4, Appli
19	7	0.8	1248	2	US-08-348-353-17	Sequence 17, Appl
20	7	0.8	1248	2	US-08-465-965-17	Sequence 17, Appl
21	7	0.8	1248	3	US-08-465-966-17	Sequence 17, Appl
22	7	0.8	1261	4	US-09-208-742-4	Sequence 4, Appli
23	7	0.8	1261	4	US-09-332-295-2	Sequence 2, Appli
24	6	0.7	6	1	US-08-093-741-80	Sequence 80, Appl
25	6	0.7	6	1	US-08-720-012-80	Sequence 80, Appl
26	6	0.7	9	5	PCT-US93-01901-17	Sequence 17, Appl
27	6	0.7	10	2	US-08-595-043A-73	Sequence 73, Appl
28	6	0.7	14	2	US-08-764-640-40	Sequence 40, Appl
29	6	0.7	14	3	US-08-973-225-40	Sequence 40, Appl
30	6	0.7	14	3	US-08-973-225-215	Sequence 215, App
31	6	0.7	14	3	US-09-244-298A-40	Sequence 40, Appl
32	6	0.7	14	4	US-09-516-704-40	Sequence 40, Appl
33	6	0.7	14	5	PCT-US95-01540-2	Sequence 2, Appli
34	6	0.7	15	2	US-08-455-625-16	Sequence 16, Appl
35	6	0.7	15	4	US-08-455-685-16	Sequence 16, Appl
36	6	0.7	15	4	US-08-060-988A-16	Sequence 16, Appl
37	6	0.7	15	5	PCT-US94-05142-16	Sequence 16, Appl
38	6	0.7	21	4	US-09-461-697-332	Sequence 332, App
39	6	0.7	23	3	US-08-851-843A-182	Sequence 182, App
40	6	0.7	23	4	US-08-974-549A-301	Sequence 301, App
41	6	0.7	23	4	US-08-854-050-182	Sequence 182, App
42	6	0.7	24	1	US-08-066-309-11	Sequence 11, Appl
43	6	0.7	24	4	US-09-220-528-70	Sequence 70, Appl
44	6	0.7	24	5	PCT-US94-05660-11	Sequence 11, Appl
45	6	0.7	27	1	US-08-140-188-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-315-794-52
; Sequence 52, Application US/09315794
; Patent No. 6197517
; GENERAL INFORMATION:
; APPLICANT:.Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATION
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-053
; CURRENT APPLICATION NUMBER: US/09/315,794
; CURRENT FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-315-794-52

Query Match 1.1%; Score 9; DB 4; Length 838;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 682 LLIHEATLE 690
|||||||||
Db 717 LLIHEATLE 725

RESULT 2
US-09-389-341-52
; Sequence 52, Application US/09389341
; Patent No. 6200803
; GENERAL INFORMATION:
; APPLICANT: Roberts, Christopher J.
; TITLE OF INVENTION: ESSENTIAL GENES OF YEAST AS TARGETS FOR ANTIFUNGAL
; TITLE OF INVENTION: AGENTS, HERBICIDES, INSECTICIDES AND ANTI-PROLIFERATIVE
; TITLE OF INVENTION: DRUGS
; FILE REFERENCE: 9301-057
; CURRENT APPLICATION NUMBER: US/09/389,341
; CURRENT FILING DATE: 1999-09-02
; EARLIER APPLICATION NUMBER: 09/315,794
; EARLIER FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 52
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae

```
US-09-389-341-52


  Query Match             1.1%;  Score 9;  DB 4;  Length 838;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     682 LLIHEATLE 690
           |||||||||
Db     717 LLIHEATLE 725


RESULT   3
US-08-930-894-5
; Sequence 5, Application US/08930894
; Patent No. 6037524
;  GENERAL INFORMATION:
;    APPLICANT:  GREENLAND, Andrew James
;    APPLICANT:  DRAPER, John
;    APPLICANT:  SKIPSEY, Marc
;    APPLICANT:  WARNER, Simon
;    TITLE OF INVENTION:  S-ADENOSYL-L-HOMOCYSTEIN HYDROLASE PROMOTER
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pillsbury Madison & Sutro
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/930,894
;      FILING DATE:  09-OCT-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/00882
;      FILING DATE:  10-APR-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9507381.3
;      FILING DATE:  10-APR-1995
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  138 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  TOBACCO SHH PROTEIN
US-08-930-894-5


  Query Match             1.0%;  Score 8;  DB 3;  Length 138;
  Best Local Similarity   100.0%;  Pred. No. 4.4;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     679 DATLLIHE 686
           ||||||||
Db      49 DATLLIHE 56


RESULT   4
US-08-981-392-33
; Sequence 33, Application US/08981392
; Patent No. 6262025
;  GENERAL INFORMATION:
;    APPLICANT:  Ish-Horowicz, David
;    APPLICANT:  Henrique, Domingos Manuel Pinto
;    APPLICANT:  Lewis, Julian Hart
;    APPLICANT:  Artavanis-Tsakonas, Spyridon
;    APPLICANT:  Gray, Grace
;    TITLE OF INVENTION:  NUCLEOTIDE AND PROTEIN SEQUENCES
;    TITLE OF INVENTION:  OF VERTEBRATE DELTA GENES AND METHODS BASED THEREON
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036/2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/981,392
;      FILING DATE:  22-DEC-1997
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Antler, Adriane M.
;      REGISTRATION NUMBER:  32,605
;      REFERENCE/DOCKET NUMBER:  7326-038
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  33:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-981-392-33


  Query Match             0.8%;  Score 7;  DB 4;  Length 40;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     203 PLSRLSP 209
           |||||||
Db       4 PLSRLSP 10


RESULT   5
US-08-685-992-7
; Sequence 7, Application US/08685992
; Patent No. 5912138
;  GENERAL INFORMATION:
;    APPLICANT:  Tonks, Nicholas
;    APPLICANT:  Flint, Andrew J.
;    TITLE OF INVENTION:  SUBSTRATE TRAPPING PROTEIN
;    TITLE OF INVENTION:  TYROSINE PHOSPHATASES
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/685,992
```

```
;      FILING DATE:  25-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  CSHL96-03
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  781-861-6240
;      TELEFAX:  781-861-9540
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  250 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-685-992-7


  Query Match             0.8%;  Score 7;  DB 2;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 77;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     783 EKRELRQ 789
           |||||||
Db     137 EKRELRQ 143


RESULT   6
US-09-144-925-7
; Sequence 7, Application US/09144925
; Patent No. 5951979
;  GENERAL INFORMATION:
;    APPLICANT:  Tonks, Nicholas
;    APPLICANT:  Flint, Andrew J.
;    TITLE OF INVENTION:  SUBSTRATE TRAPPING PROTEIN
;    TITLE OF INVENTION:  TYROSINE PHOSPHATASES
;    NUMBER OF SEQUENCES:  36
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02421-4799
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/144,925
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/685,992
;      FILING DATE:  July 25, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  CSHL96-03Z
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  781-861-6240
;      TELEFAX:  781-861-9540
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  250 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-09-144-925-7


  Query Match             0.8%;  Score 7;  DB 2;  Length 250;
  Best Local Similarity   100.0%;  Pred. No. 77;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     783 EKRELRQ 789
           |||||||
Db     137 EKRELRQ 143


RESULT   7
US-08-036-210-13
; Sequence 13, Application US/08036210
; Patent No. 5585233
;  GENERAL INFORMATION:
;    APPLICANT:  Moller, Niels P.H.
;    APPLICANT:  Moller, Karin B.
;    APPLICANT:  Ullrich, Axel
;    TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/036,210
;      FILING DATE:  23-MAR-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7683-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  289 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-036-210-13


  Query Match             0.8%;  Score 7;  DB 1;  Length 289;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     783 EKRELRQ 789
           |||||||
Db     177 EKRELRQ 183


RESULT   8
US-08-449-609-13
; Sequence 13, Application US/08449609
; Patent No. 5952212
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Moller, Niels P.H.
;    APPLICANT:  Moller, Karin B.
;    APPLICANT:  Ullrich, Axel
;    TITLE OF INVENTION:  PTP-S31:  A NOVEL PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PENNIE & EDMONDS
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036-2711
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/449,609
;      FILING DATE:  24-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/036,210
;      FILING DATE:  23-MAR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Misrock, S. Leslie
;      REGISTRATION NUMBER:  18,872
;      REFERENCE/DOCKET NUMBER:  7683-025
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-8864/9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  289 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-449-609-13


  Query Match             0.8%;  Score 7;  DB 2;  Length 289;
  Best Local Similarity   100.0%;  Pred. No. 88;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     783 EKRELRQ 789
           |||||||
Db     177 EKRELRQ 183


RESULT  9
US-08-878-989-7
; Sequence 7, Application US/08878989
; Patent No. 5885803
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl G.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  DISEASE ASSOCIATED PROTEIN
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,989
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0321 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  328 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  PENITUT01
;      CLONE:  1452972
US-08-878-989-7


  Query Match             0.8%;  Score 7;  DB 2;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     503 LVNISPD 509
           |||||||
Db     120 LVNISPD 126


RESULT  10
US-09-272-796-7
; Sequence 7, Application US/09272796
; Patent No. 6207148
;  GENERAL INFORMATION:
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Hillman, Jennifer L.
;    APPLICANT:  Corley, Neil C.
;    APPLICANT:  Guegler, Karl G.
;    APPLICANT:  Lal, Preeti
;    APPLICANT:  Goli, Surya K.
;    APPLICANT:  Shah, Purvi
;    TITLE OF INVENTION:  DISEASE ASSOCIATED PROTEIN
;    TITLE OF INVENTION:  KINASES
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/272,796
;      FILING DATE:
```

```
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/878,989
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0321 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  328 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  PENITUT01
;      CLONE:  1452972
US-09-272-796-7


  Query Match             0.8%;  Score 7;  DB 4;  Length 328;
  Best Local Similarity   100.0%;  Pred. No. 99;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     503 LVNISPD 509
           |||||||
Db     120 LVNISPD 126


RESULT  11
US-08-677-049-8
; Sequence 8, Application US/08677049
; Patent No. 5858707
;  GENERAL INFORMATION:
;    APPLICANT:  Guimaraes, M. Jorge
;    APPLICANT:  Bazan, J. Fernando
;    APPLICANT:  McClanahan, Terrill K.
;    APPLICANT:  Zlotnik, Albert
;    TITLE OF INVENTION:  PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
;    TITLE OF INVENTION:  NUCLEIC ACIDS; ANTIBODIES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/677,049
;      FILING DATE:  03-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/000,788
;      FILING DATE:  03-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0511
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-852-9196
;      TELEFAX:  415-496-1200
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  432 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  115..144
;      OTHER INFORMATION:  /note= "Encompasses TM 4 of Figure
;      OTHER INFORMATION:  4"
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  325..359
;      OTHER INFORMATION:  /note= "Encompasses TM 9 of Figure
;      OTHER INFORMATION:  4"
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  363..390
;      OTHER INFORMATION:  /note= "Encompasses TM 10 of Figure
;      OTHER INFORMATION:  4"
US-08-677-049-8


  Query Match             0.8%;  Score 7;  DB 2;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     271 IAPIIAA 277
           |||||||
Db      76 IAPIIAA 82


RESULT  12
US-08-677-049-10
; Sequence 10, Application US/08677049
; Patent No. 5858707
;  GENERAL INFORMATION:
;    APPLICANT:  Guimaraes, M. Jorge
;    APPLICANT:  Bazan, J. Fernando
;    APPLICANT:  McClanahan, Terrill K.
;    APPLICANT:  Zlotnik, Albert
;    TITLE OF INVENTION:  PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
;    TITLE OF INVENTION:  NUCLEIC ACIDS; ANTIBODIES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DNAX Research Institute
;      STREET:  901 California Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94304-1104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/677,049
;      FILING DATE:  03-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/000,788
;      FILING DATE:  03-JUL-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ching, Edwin P.
;      REGISTRATION NUMBER:  34,090
;      REFERENCE/DOCKET NUMBER:  DX0511
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-852-9196
;      TELEFAX:  415-496-1200
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  460 amino acids
```

```
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  117..149
;      OTHER INFORMATION:  /note= "Encompasses TM 4 of Figure
;      OTHER INFORMATION:  4"
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  328..362
;      OTHER INFORMATION:  /note= "Encompasses TM 9 of Figure
;      OTHER INFORMATION:  4"
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  365..392
;      OTHER INFORMATION:  /note= "Encompasses TM 10 of Figure
;      OTHER INFORMATION:  4"
US-08-677-049-10


  Query Match             0.8%;  Score 7;  DB 2;  Length 460;
  Best Local Similarity   100.0%;  Pred. No. 1.3e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     535 VLGTLAA 541
           |||||||
Db     410 VLGTLAA 416


RESULT  13
US-08-320-559-31
; Sequence 31, Application US/08320559
; Patent No. 5633135
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods for
;    TITLE OF INVENTION:  Detection and Treatment of Acute Leukemias
;    TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the
;    TITLE OF INVENTION:  All-1 Region
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  PC-DOS
;      SOFTWARE:  WORDPERFECT 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/320,559
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/062,443
;      FILING DATE:  14 MAY 1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/971,094
;      FILING DATE:  30-OCT-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/888,830
;      FILING DATE:  27-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/805,093
;      FILING DATE:  11-DEC-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-0855
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  559 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-320-559-31


  Query Match             0.8%;  Score 7;  DB 1;  Length 559;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      29 RERPRKD 35
           |||||||
Db     202 RERPRKD 208


RESULT  14
US-08-545-860D-31
; Sequence 31, Application US/08545860D
; Patent No. 6040140
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods
;    TITLE OF INVENTION:  for Detection and Treatment of Acute Leukemias
;    TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the All-1 Region
;    NUMBER OF SEQUENCES:  94
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewitz &
;      ADDRESSEE:  No. 6040140ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/545,860D
;      FILING DATE:  07-MAR-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04496
;      FILING DATE:  22-APR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/10930
;      FILING DATE:  09-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/327,392
;      FILING DATE:  19-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/320,559
;      FILING DATE:  11-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/062,443
;      FILING DATE:  14-MAY-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,094
;      FILING DATE:  30-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/888,839
;      FILING DATE:  27-MAY-1992
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/805,093
;      FILING DATE:  11-DEC-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca Esq., Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-1262
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  559 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-545-860D-31


  Query Match             0.8%;  Score 7;  DB 3;  Length 559;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      29 RERPRKD 35
           |||||||
Db     202 RERPRKD 208


RESULT  15
PCT-US94-04496-31
; Sequence 31, Application PC/TUS9404496
;  GENERAL INFORMATION:
;    APPLICANT:  Croce, Carlo
;    APPLICANT:  Canaani, Eli
;    TITLE OF INVENTION:  Diagnostics, Therapeutics and Methods
;    TITLE OF INVENTION:  for Detection and Treatment of Acute Leukemias
;    TITLE OF INVENTION:  Resulting from Chromosome Abnormalities in the All-1
;    NUMBER OF SEQUENCES:  86
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewitz &
;      ADDRESSEE:  Norris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  Pennsylvania
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/04496
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca Esq., Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  TJU-1242
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  559 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-04496-31


  Query Match             0.8%;  Score 7;  DB 5;  Length 559;
  Best Local Similarity   100.0%;  Pred. No. 1.6e+02;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      29 RERPRKD 35
           |||||||
Db     202 RERPRKD 208


Search completed: March 29, 2002, 10:23:12
Job time: 139 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 29, 2002, 10:21:43 ; Search time 32.11 Seconds
(without alignments)
3762.720 Million cell updates/sec

Title: US-09-988-687-2
Perfect score: 826
Sequence: 1 MWALCSLLRSAAGRTMSQGR.....EPQQKRAHTEEPQAKKVRAQ 826

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	826	100.0	826	4	Q9HAS8	Q9has8 homo sapien
2	725	87.8	826	4	Q9BQ52	Q9bq52 homo sapien
3	430	52.1	854	4	Q9NVT1	Q9nvt1 homo sapien
4	254	30.8	826	6	Q9GL72	Q9gl72 pan troglod
5	157	19.0	826	6	Q9GL73	Q9gl73 gorilla gor
6	51	6.2	824	11	Q99MF1	Q99mf1 mus musculu
7	51	6.2	831	11	Q9EPZ2	Q9epz2 mus musculu
8	51	6.2	831	11	Q99MF0	Q99mf0 mus musculu
9	46	5.6	46	4	Q9HAS9	Q9has9 homo sapien
10	39	4.7	39	4	Q9HAS7	Q9has7 homo sapien
11	34	4.1	225	11	Q9D1A8	Q9d1a8 mus musculu
12	34	4.1	435	11	Q9CTA2	Q9cta2 mus musculu
13	10	1.2	805	10	Q9LWQ4	Q9lwq4 oryza sativ
14	10	1.2	837	10	Q9M819	Q9m819 arabidopsis
15	9	1.1	743	5	Q9V5J4	Q9v5j4 drosophila
16	9	1.1	789	5	Q9GZ73	Q9gz73 drosophila
17	8	1.0	121	11	Q9CXB1	Q9cxb1 mus musculu
18	8	1.0	147	1	Q9V017	Q9v017 pyrococcus
19	8	1.0	171	10	Q9SP98	Q9sp98 solanum cha
20	8	1.0	216	5	Q21169	Q21169 caenorhabdi
21	8	1.0	255	2	O86707	O86707 streptomyce
22	8	1.0	260	1	Q9V0U5	Q9v0u5 pyrococcus
23	8	1.0	280	1	O59139	O59139 pyrococcus
24	8	1.0	307	1	O58883	O58883 pyrococcus
25	8	1.0	312	10	Q9SDP1	Q9sdp1 allium cepa
26	8	1.0	360	4	Q9BUM1	Q9bum1 homo sapien
27	8	1.0	363	4	Q9NS99	Q9ns99 homo sapien
28	8	1.0	363	4	Q9H777	Q9h777 homo sapien
29	8	1.0	388	2	O31265	O31265 arthrobacte
30	8	1.0	445	5	Q9TZZ6	Q9tzz6 trypanosoma
31	8	1.0	450	10	Q42939	Q42939 nicotiana s
32	8	1.0	473	2	Q9RB44	Q9rb44 clostridium
33	8	1.0	485	10	Q9SWF5	Q9swf5 lycopersico
34	8	1.0	485	10	Q9LK36	Q9lk36 arabidopsis
35	8	1.0	562	5	Q9BHH9	Q9bhh9 leishmania
36	8	1.0	643	2	Q9K6I7	Q9k6i7 bacillus ha
37	8	1.0	806	2	Q9XAD8	Q9xad8 streptomyce
38	8	1.0	863	4	Q9UPV4	Q9upv4 homo sapien
39	8	1.0	876	5	Q9VEA2	Q9vea2 drosophila
40	8	1.0	927	10	Q9LU20	Q9lu20 arabidopsis
41	8	1.0	989	2	Q9HUR6	Q9hur6 pseudomonas
42	8	1.0	1630	4	Q9Y4D8	Q9y4d8 homo sapien
43	7	0.8	54	3	Q00378	Q00378 coccidioide
44	7	0.8	63	2	Q50125	Q50125 mycobacteri
45	7	0.8	71	8	Q9MUR2	Q9mur2 mesostigma

ALIGNMENTS

RESULT 1
Q9HAS8
ID Q9HAS8 PRELIMINARY; PRT; 826 AA.
AC Q9HAS8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PUTATIVE PROSTATE CANCER SUSCEPTIBILITY PROTEIN HPC2/ELAC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tavtigian S.V., Simard J., Teng D.H.F., Baumgard M., Beck A.,
RA Camp N.J., Carillo A.R., Chen Y., Dayananth P., Desrochers M.,
RA Dumont M., Farnham J.M., Frank D., Frye C., Ghaffari S., Gupte J.S.,
RA Hu R., Iliev D., Janecki T., Kort E.N., Laity K.E., Leavitt A.,
RA Leblanc G., McArthur-Morrison J., Pederson A., Penn B., Peterson K.T.,
RA Reid J.E., Richards S., Schroeder M., Smith R., Snyder S.C.,
RA Swedlund B., Swensen J., Thomas A., Tranchant M., Woodland A.-M.,
RA Labrie F., Skolnick M.H., Neuhausen S., Rommens J.,
RA Cannon-Albright L.A.;
RT "A strong candidate prostate cancer susceptibility gene at chromosome
RT 17p.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF304370; AAG24441.1; -.
SQ SEQUENCE 826 AA; 92218 MW; 4AE701C755EC7339 CRC64;

Query Match 100.0%; Score 826; DB 4; Length 826;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 826; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60

Qy      61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120
```

```
Qy    121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180

Qy    181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240

Qy    241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300

Qy    301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360

Qy    361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420

Qy    421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480

Qy    481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540

Qy    541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600

Qy    601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660

Qy    661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720

Qy    721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEE 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEE 780

Qy    781 RREKRELRQVRAALLSRELAGGLEDGEPQQKRAHTEEPQAKKVRAQ 826
          |||||||||||||||||||||||||||||||||||||||||||||
Db    781 RREKRELRQVRAALLSRELAGGLEDGEPQQKRAHTEEPQAKKVRAQ 826
```

```
RESULT   2
Q9BQ52
ID   Q9BQ52      PRELIMINARY;      PRT;   826 AA.
AC   Q9BQ52;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE PROSTATE CANCER SUSCEPTIBILITY PROTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG CARCINOMA;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG CARCINOMA;
RA   Strausberg R.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC004158; AAH04158.1; -.
DR   EMBL; BC001939; AAH01939.1; -.
```

```
SQ   SEQUENCE   826 AA;  92245 MW;  8B3A38C355757AAE CRC64;


  Query Match             87.8%;  Score 725;  DB 4;  Length 826;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches  825;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60

Qy     61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120

Qy    121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180

Qy    181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSELNENEPHLPHGVSQRRGVRDSSLV 240

Qy    241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300

Qy    301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360

Qy    361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420

Qy    421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480

Qy    481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540

Qy    541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600

Qy    601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660

Qy    661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720

Qy    721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEE 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEE 780

Qy    781 RREKRELRQVRAALLSRELAGGLEDGEPQQKRAHTEEPQAKKVRAQ 826
          |||||||||||||||||||||||||||||||||||||||||||||
Db    781 RREKRELRQVRAALLSRELAGGLEDGEPQQKRAHTEEPQAKKVRAQ 826
```

```
RESULT   3
Q9NVT1
ID   Q9NVT1      PRELIMINARY;      PRT;   854 AA.
AC   Q9NVT1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   CDNA FLJ10530 FIS, CLONE NT2RP2000985.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK001392; BAA91666.1; -.
SQ   SEQUENCE   854 AA;  95138 MW;  DF9F32846C7166B1 CRC64;


  Query Match             52.1%;  Score 430;  DB 4;  Length 854;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches  750;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60

Qy      61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120

Qy     121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAMRPHSAPEYEDETMTV 180

Qy     181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240

Qy     241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300

Qy     301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360

Qy     361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRC-KEGPTLSVPMVQGE 419

Qy     421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 479

Qy     481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 539

Qy     541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600
           |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     540 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYYNQCQEVLH 599

Qy     601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 659

Qy     661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 719

Qy     721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVC 753
           |||||||||||||||||||||||||||||||||
Db     720 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVC 752


RESULT   4
Q9GL72
ID   Q9GL72      PRELIMINARY;      PRT;   826 AA.
AC   Q9GL72;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ELAC2.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tavtigian S.V., Simard J., Teng D.H.F., Baumgard M., Dayananth P.,
RA   Desrochers M., Dumont M., Farnham J.M., Frank D., Frye C.,
RA   Ghaffari S., Gupte J.S., Hu R., Iliev D., Janecki T., Kort E.N.,
RA   Laity K., Leavitt A., Leblanc G., McArthur-Morrison J., Pederson A.,
RA   Penn B., Peterson K.T., Reid J.E., Richards S., Schroeder M.,
RA   Smith R., Snyder S.C., Swedlund B., Swensen J., Thomas A.,
RA   Tranchant M., Woodland A.-M., Labrie F., Skolnick M.H., Neuhausen S.,
RA   Rommens J., Cannon-Albright L.A.;
RT   "Pan troglodytes ortholog of human HPC2/ELAC2.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF308698; AAG24920.1; -.
SQ   SEQUENCE   826 AA;  92324 MW;  51A62814155E7191 CRC64;


  Query Match             30.8%;  Score 254;  DB 6;  Length 826;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-253;
  Matches  254;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      92 QRLMQEHKLKVARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAI 151
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92 QRLMQEHKLKVARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAI 151

Qy     152 KIFSGPLKGIELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPER 211
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152 KIFSGPLKGIELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPER 211

Qy     212 SSDSESNENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAI 271
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212 SSDSESNENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAI 271

Qy     272 APIIAAVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQG 331
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     272 APIIAAVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQG 331

Qy     332 KADAPVALVVHMAP 345
           ||||||||||||||
Db     332 KADAPVALVVHMAP 345


RESULT   5
Q9GL73
ID   Q9GL73      PRELIMINARY;      PRT;   826 AA.
AC   Q9GL73;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ELAC2.
GN   ELAC2.
OS   Gorilla gorilla (gorilla).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX   NCBI_TaxID=9593;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tavtigian S.V., Simard J., Teng D.H.F., Baumgard M., Beck A.,
RA   Camp N.J., Carillo A.R., Chen Y., Dayananth P., Desrochers M.,
RA   Dumont M., Farnham J.M., Frank D., Frye C., Ghaffari S., Gupte J.S.,
```

```
RA   Hu R., Iliev D., Janecki T., Kort E.N., Laity K.E., Leavitt A.,
RA   Leblanc G., McArthur-Morrison J., Pederson A., Penn B., Peterson K.T.,
RA   Reid J.E., Richards S., Schroeder M., Smith R., Snyder S.C.,
RA   Swedlund B., Swensen J., Thomas A., Tranchant M., Woodland A.-M.,
RA   Labrie F., Skolnick M.H., Neuhausen S., Rommens J.,
RA   Cannon-Albright L.A.;
RT   "Gorilla gorilla ortholog of human HPC2/ELAC2.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF308694; AAG24916.1; -.
SQ   SEQUENCE   826 AA;  92260 MW;  D323B5F3D8B294A6 CRC64;


  Query Match           19.0%;  Score 157;  DB 6;  Length 826;
  Best Local Similarity  100.0%;  Pred. No. 4.3e-153;
  Matches  157;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     583 QLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLV 642
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     583 QLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLV 642

Qy     643 RHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTH 702
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     643 RHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTH 702

Qy     703 STTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNF 739
           |||||||||||||||||||||||||||||||||||||
Db     703 STTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNF 739


RESULT   6
Q99MF1
ID   Q99MF1      PRELIMINARY;      PRT;   824 AA.
AC   Q99MF1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE PROSTATE CANCER SUSCEPTIBILITY PROTEIN.
GN   ELAC2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Frank D.C., Swedlund B., Dumont M., Tavtigian S.V., Simard J.,
RA   Teng D.H.F., Baumgard M., Beck A., Camp N.J., Carillo A.R., Chen Y.,
RA   Dayananth P., Desrochers M., Farnham J.M., Frye C., Ghaffari S.,
RA   Gupte J.S., Hu R., Iliev D., Janecki T., Kort E.N., Laity K.E.,
RA   Leavitt A., Leblanc G., McArthur-Morrison J., Pederson A., Penn B.,
RA   Peterson K.T., Reid J.E., Richards S., Schroeder M., Smith R.,
RA   Snyder S.C., Swensen J., Thomas A., Tranchant M., Woodland A.-M.,
RA   Labrie F., Skolnick M.H., Neuhausen S., Rommens J.,
RA   Cannon-Albright L.A.;
RT   "Mouse Elac2-containing genomic DNA.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF348157; AAK29421.1; -.
SQ   SEQUENCE   824 AA;  92085 MW;  77F870F0E27D6746 CRC64;


  Query Match            6.2%;  Score 51;  DB 11;  Length 824;
  Best Local Similarity  100.0%;  Pred. No. 2.1e-43;
  Matches   51;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     123 GMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEY 173
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 GMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEY 165


RESULT   7
Q9EPZ2
ID   Q9EPZ2      PRELIMINARY;      PRT;   831 AA.
AC   Q9EPZ2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   ELAC2.
GN   1110017O07RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tavtigian S.V., Simard J., Teng D.H.F., Baumgard M., Beck A.,
RA   Camp N.J., Carillo A.R., Chen Y., Dayananth P., Desrochers M.,
RA   Dumont M., Farnham J.M., Frank D., Frye C., Ghaffari S., Gupte J.S.,
RA   Hu R., Iliev D., Janecki T., Kort E.N., Laity K.E., Leavitt A.,
RA   Leblanc G., McArthur-Morrison J., Pederson A., Penn B., Peterson K.T.,
RA   Reid J.E., Richards S., Schroeder M., Smith R., Snyder S.C.,
RA   Swedlund B., Swensen J., Thomas A., Tranchant M., Woodland A.-M.,
RA   Labrie F., Skolnick M.H., Neuhausen S., Rommens J.,
RA   Cannon-Albright L.A.;
RT   "Mouse ortholog of human HPC2/ELAC2.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF308696; AAG24918.2; -.
DR   MGD; MGI:1915876; 1110017O07Rik.
SQ   SEQUENCE   831 AA;  92631 MW;  4EA7DFAC292E2B32 CRC64;


  Query Match            6.2%;  Score 51;  DB 11;  Length 831;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-43;
  Matches   51;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     123 GMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEY 173
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 GMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEY 165


RESULT   8
Q99MF0
ID   Q99MF0      PRELIMINARY;      PRT;   831 AA.
AC   Q99MF0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PUTATIVE PROSTATE CANCER SUSCEPTIBILITY PROTEIN.
GN   ELAC2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Frank D.C., Swedlund B., Dumont M., Tavtigian S.V., Simard J.,
RA   Teng D.H.F., Baumgard M., Beck A., Camp N.J., Carillo A.R., Chen Y.,
RA   Dayananth P., Desrochers M., Farnham J.M., Frye C., Ghaffari S.,
RA   Gupte J.S., Hu R., Iliev D., Janecki T., Kort E.N., Laity K.E.,
RA   Leavitt A., Leblanc G., McArthur-Morrison J., Pederson A., Penn B.,
RA   Peterson K.T., Reid J.E., Richards S., Schroeder M., Smith R.,
RA   Snyder S.C., Swensen J., Thomas A., Tranchant M., Woodland A.-M.,
RA   Labrie F., Skolnick M.H., Neuhausen S., Rommens J.,
RA   Cannon-Albright L.A.;
RT   "Mouse Elac2-containing genomic DNA.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF348157; AAK29420.1; -.
SQ   SEQUENCE   831 AA;  92729 MW;  F49FC6C65C9FDDEC CRC64;


  Query Match            6.2%;  Score 51;  DB 11;  Length 831;
  Best Local Similarity  100.0%;  Pred. No. 2.2e-43;
  Matches   51;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     123 GMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEY 173
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     115 GMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEY 165
```

```
RESULT   9
Q9HAS9
ID   Q9HAS9      PRELIMINARY;      PRT;    46 AA.
AC   Q9HAS9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE PROSTATE CANCER SUSCEPTIBILITY PROTEIN HPC2/ELAC2 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tavtigian S.V., Simard J., Teng D.H.F., Baumgard M., Beck A.,
RA   Camp N.J., Carillo A.R., Chen Y., Dayananth P., Desrochers M.,
RA   Dumont M., Farnham J.M., Frank D., Frye C., Ghaffari S., Gupte J.S.,
RA   Hu R., Iliev D., Janecki T., Kort E.N., Laity K.E., Leavitt A.,
RA   Leblanc G., McArthur-Morrison J., Pederson A., Penn B., Peterson K.T.,
RA   Reid J.E., Richards S., Schroeder M., Smith R., Snyder S.C.,
RA   Swedlund B., Swensen J., Thomas A., Tranchant M., Woodland A.-M.,
RA   Labrie F., Skolnick M.H., Neuhausen S., Rommens J.,
RA   Cannon-Albright L.A.;
RT   "A strong candidate prostate cancer susceptibility gene at chromosome
RT   17p.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF304369; AAG24440.1; -.
FT   NON_TER       1      1
FT   NON_TER      46     46
SQ   SEQUENCE   46 AA;  5004 MW;  59C8CDF4E5FDDBBE CRC64;


  Query Match             5.6%;  Score 46;  DB 4;  Length 46;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-39;
  Matches   46;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     508 PDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHT 553
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 PDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHT 46


RESULT  10
Q9HAS7
ID   Q9HAS7      PRELIMINARY;      PRT;    39 AA.
AC   Q9HAS7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE PROSTATE CANCER SUSCEPTIBILITY PROTEIN HPC2/ELAC2 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tavtigian S.V., Simard J., Teng D.H.F., Baumgard M., Beck A.,
RA   Camp N.J., Carillo A.R., Chen Y., Dayananth P., Desrochers M.,
RA   Dumont M., Farnham J.M., Frank D., Frye C., Ghaffari S., Gupte J.S.,
RA   Hu R., Iliev D., Janecki T., Kort E.N., Laity K.E., Leavitt A.,
RA   Leblanc G., McArthur-Morrison J., Pederson A., Penn B., Peterson K.T.,
RA   Reid J.E., Richards S., Schroeder M., Smith R., Snyder S.C.,
RA   Swedlund B., Swensen J., Thomas A., Tranchant M., Woodland A.-M.,
RA   Labrie F., Skolnick M.H., Neuhausen S., Rommens J.,
RA   Cannon-Albright L.A.;
RT   "A strong candidate prostate cancer susceptibility gene at chromosome
RT   17p.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF304371; AAG24439.1; -.
FT   NON_TER       1      1
FT   NON_TER      39     39
SQ   SEQUENCE   39 AA;  4616 MW;  F275E87B34DB9688 CRC64;


  Query Match             4.7%;  Score 39;  DB 4;  Length 39;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-32;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     188 EQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPH 226
           |||||||||||||||||||||||||||||||||||||||||
Db       1 EQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPH 39


RESULT  11
Q9D1A8
ID   Q9D1A8      PRELIMINARY;      PRT;   225 AA.
AC   Q9D1A8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   1110017O07RIK PROTEIN.
GN   1110017O07RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK003759; BAB22981.1; -.
DR   MGD; MGI:1915876; 1110017O07Rik.
SQ   SEQUENCE   225 AA;  25346 MW;  A82E61B36A57DC60 CRC64;


  Query Match             4.1%;  Score 34;  DB 11;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-26;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     676 MGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAI 709
           ||||||||||||||||||||||||||||||||||
Db      73 MGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAI 106


RESULT  12
Q9CTA2
ID   Q9CTA2      PRELIMINARY;      PRT;   435 AA.
AC   Q9CTA2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   DNA SEGMENT, CHR 11, WAYNE STATE UNIVERSITY 80, EXPRESSED (FRAGMENT).
GN   D11WSU80E.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004136; BAB23185.1; -.
DR   MGD; MGI:106352; D11Wsu80e.
FT   NON_TER       1       1
SQ   SEQUENCE   435 AA;  48949 MW;  4730D4D49524CB29 CRC64;


  Query Match             4.1%;  Score 34;  DB 11;  Length 435;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-26;
  Matches   34;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     676 MGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAI 709
           ||||||||||||||||||||||||||||||||||
Db     276 MGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAI 309


RESULT  13
Q9LWQ4
ID   Q9LWQ4      PRELIMINARY;      PRT;   805 AA.
AC   Q9LWQ4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   SIMILAR TO SCHIZOSACCHAROMYCES POMBE CHROMOSOME I COSMID C1D4.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0431F01.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001550; BAA92999.1; -.
SQ   SEQUENCE   805 AA;  89046 MW;  56F8366CB2FF7E0C CRC64;


  Query Match             1.2%;  Score 10;  DB 10;  Length 805;
  Best Local Similarity   100.0%;  Pred. No. 0.56;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     658 GWKVVYSGDT 667
           ||||||||||
Db     675 GWKVVYSGDT 684


RESULT  14
Q9M819
ID   Q9M819      PRELIMINARY;      PRT;   837 AA.
AC   Q9M819;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   F9I5.1 PROTEIN.
GN   F9I5.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RA   Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
RA   Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA   Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT   "Arabidopsis thaliana chromosome I BAC F9I5 genomic sequence.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC022354; AAF29402.1; -.
SQ   SEQUENCE   837 AA;  92978 MW;  4B5BFDF3F3B698CE CRC64;


  Query Match             1.2%;  Score 10;  DB 10;  Length 837;
  Best Local Similarity   100.0%;  Pred. No. 0.58;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     511 SLLLDCGEGT 520
           ||||||||||
Db     515 SLLLDCGEGT 524


RESULT  15
Q9V5J4
ID   Q9V5J4      PRELIMINARY;      PRT;   743 AA.
AC   Q9V5J4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CG3298 PROTEIN.
GN   JHI-1 OR CG3298.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
```

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003830; AAF58812.1; -.
DR FlyBase; FBgn0028426; JhI-1.
SQ SEQUENCE 743 AA; 82895 MW; 704327777345D01D CRC64;

Query Match 1.1%; Score 9; DB 5; Length 743;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      82 RYLFNCGEG 90
           |||||||||
Db      60 RYLFNCGEG 68
```

Search completed: March 29, 2002, 10:24:21
Job time: 158 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 29, 2002, 10:22:23 ; Search time 15.35 Seconds
(without alignments)
1972.974 Million cell updates/sec

Title: US-09-988-687-2
Perfect score: 826
Sequence: 1 MWALCSLLRSAAGRTMSQGR.....EPQQKRAHTEEPQAKKVRAQ 826

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11	1.3	809	1	YATA_SCHPO	Q10155 schizosacch
2	9	1.1	838	1	YK59_YEAST	P36159 saccharomyc
3	8	1.0	423	1	DCOR_TRYBB	P07805 trypanosoma
4	8	1.0	485	1	SAHH_ARATH	O23255 arabidopsis
5	8	1.0	485	1	SAHH_CATRO	P35007 catharanthu
6	8	1.0	485	1	SAHH_MESCR	P93253 mesembryant
7	8	1.0	485	1	SAHH_PETCR	Q01781 petroselinu
8	8	1.0	485	1	SAHH_PHASS	P50249 phalaenopsi
9	8	1.0	485	1	SAHH_TOBAC	P50248 nicotiana t
10	8	1.0	485	1	SAHH_WHEAT	P32112 triticum ae
11	8	1.0	836	1	Y167_HUMAN	Q99490 homo sapien
12	8	1.0	1276	1	MDR2_CRIGR	P21449 cricetulus
13	7	0.8	71	1	YCX4_MESVI	Q9mur2 mesostigma
14	7	0.8	129	1	IR08_HCMVA	P16806 human cytom
15	7	0.8	129	1	VG42_HSVI1	Q00101 ictalurid h
16	7	0.8	211	1	LEF7_NPVOP	O10362 orgyia pseu
17	7	0.8	216	1	INAA_ECOLI	P27294 escherichia
18	7	0.8	220	1	TPIS_CHLAU	P96744 chloroflexu
19	7	0.8	241	1	KORA_STRLI	P22405 streptomyce
20	7	0.8	258	1	SN29_HUMAN	O95721 homo sapien
21	7	0.8	259	1	PPNK_MYCPN	P75508 mycoplasma
22	7	0.8	264	1	RPOD_ARCFU	O28002 archaeoglob
23	7	0.8	264	1	SNAI_MOUSE	Q02085 mus musculu
24	7	0.8	264	1	YDCV_ECOLI	P77505 escherichia
25	7	0.8	273	1	YD39_MYCTU	Q10648 mycobacteri
26	7	0.8	281	1	YHIR_HAEIN	P31777 haemophilus
27	7	0.8	284	1	BCHO_RHOCA	P26174 rhodobacter
28	7	0.8	284	1	YD39_MYCLE	P50474 mycobacteri
29	7	0.8	287	1	YKFA_ECOLI	P75678 escherichia
30	7	0.8	297	1	YPGA_PORGI	Q51834 porphyromon
31	7	0.8	303	1	GDFF_RAT	Q9z0j6 rattus norv
32	7	0.8	326	1	LEU3_ARCFU	O29627 archaeoglob
33	7	0.8	338	1	G3P_SCHMA	P20287 schistosoma
34	7	0.8	343	1	GPDA_RHILO	P58142 rhizobium l
35	7	0.8	394	1	FXD3_CHICK	P79772 gallus gall
36	7	0.8	427	1	AFC2_ARATH	P51567 arabidopsis
37	7	0.8	432	1	CXA8_HUMAN	P48165 homo sapien
38	7	0.8	432	1	PYRP_BACCL	P41006 bacillus ca
39	7	0.8	435	1	PROA_AQUAE	O67166 aquifex aeo
40	7	0.8	439	1	CXA8_MOUSE	P28236 mus musculu
41	7	0.8	439	1	CXA8_SHEEP	P55917 ovis aries
42	7	0.8	444	1	KRAC_DICDI	P54644 dictyosteli
43	7	0.8	460	1	YWDJ_BACSU	P39618 bacillus su
44	7	0.8	461	1	NORM_VIBCH	Q9kru4 vibrio chol
45	7	0.8	463	1	YDI4_SCHPO	Q92342 schizosacch

ALIGNMENTS

RESULT 1
YATA_SCHPO
ID YATA_SCHPO STANDARD; PRT; 809 AA.
AC Q10155;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 90.6 KDA PROTEIN C1D4.10 IN CHROMOSOME I.
GN SPAC1D4.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: TO S.POMBE SPBC3D5.03C AND SOME, TO YEAST YKR079C.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; Z69239; CAA93219.1; -.
DR HSSP; P15039; 1PRV.
KW Hypothetical protein.
SQ SEQUENCE 809 AA; 90602 MW; 350FBE7B05FBF880 CRC64;

Query Match 1.3%; Score 11; DB 1; Length 809;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 681 TLLIHEATLED 691
||||||||||||
Db 704 TLLIHEATLED 714

RESULT 2
YK59_YEAST
ID YK59_YEAST STANDARD; PRT; 838 AA.
AC P36159;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOTHETICAL 96.8 KDA PROTEIN IN SIS2-MTD1 INTERGENIC REGION.
GN YKR079C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

```
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pohl T.M., Pohl F.M.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: SOME, TO S.POMBE SPAC1D4.10.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z28304; CAA82158.1; -.
DR   PIR; S38156; S38156.
DR   SGD; S0001787; YKR079C.
KW   Hypothetical protein.
SQ   SEQUENCE   838 AA;  96816 MW;  473A69D9C10167AB CRC64;


  Query Match             1.1%;  Score 9;  DB 1;  Length 838;
  Best Local Similarity   100.0%;  Pred. No. 1.3;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    682 LLIHEATLE 690
          |||||||||
Db    717 LLIHEATLE 725


RESULT   3
DCOR_TRYBB
ID   DCOR_TRYBB     STANDARD;      PRT;   423 AA.
AC   P07805;
DT   01-AUG-1988 (Rel. 08, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ORNITHINE DECARBOXYLASE (EC 4.1.1.17) (ODC).
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87250494; PubMed=3036823;
RA   Phillips M.A., Coffino P., Wang C.C.;
RT   "Cloning and sequencing of the ornithine decarboxylase gene from
RT   Trypanosoma brucei. Implications for enzyme turnover and selective
RT   difluoromethylornithine inhibition.";
RL   J. Biol. Chem. 262:8721-8727(1987).
CC   -!- CATALYTIC ACTIVITY: L-ORNITHINE = PUTRESCINE + CO(2).
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: FIRST STEP AND ALSO THE RATE-LIMITING STEP IN THE PATHWAY
CC       OF POLYAMINE BIOSYNTHESIS.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF ORNITHINE, DAP, AND ARGININE
CC       DECARBOXYLASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02771; AAA30218.1; ALT_INIT.
DR   EMBL; J02771; AAA30219.1; -.
DR   PIR; A29309; DCUTOB.
DR   InterPro; IPR000183; Orn_DAP_Arg_decarbxylse.
DR   Pfam; PF00278; Orn_DAP_Arg_deC; 2.
DR   PRINTS; PR01179; ODADCRBXLASE.
DR   PRINTS; PR01182; ORNDCRBXLASE.
DR   PROSITE; PS00878; ODR_DC_2_1; 1.
DR   PROSITE; PS00879; ODR_DC_2_2; 2.
KW   Lyase; Decarboxylase; Pyridoxal phosphate; Polyamine biosynthesis.
FT   BINDING      67     67       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   ACT_SITE    358    358       BY SIMILARITY.
SQ   SEQUENCE   423 AA;  46881 MW;  924A5AA6C4CD2C36 CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 423;
  Best Local Similarity   100.0%;  Pred. No. 7.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    534 RVLGTLAA 541
          ||||||||
Db     73 RVLGTLAA 80


RESULT   4
SAHH_ARATH
ID   SAHH_ARATH     STANDARD;      PRT;   485 AA.
AC   O23255; O81847;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE
DE   HYDROLASE) (ADOHCYASE).
GN   SAHH OR AT4G13940 OR DL3010W.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA;
RA   Belbahri L., Elleuch H., Villarroel R., Inze D., Thomas D.,
RA   Thomasset B.;
RT   "The isolation of an Arabidopsis thaliana cDNA clone encoding S-
RT   adenosyl-L-homocysteine hydrolase.";
RL   (In) Plant Gene Register PGR99-139.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=98121113; PubMed=9461215;
RA   Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA   Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA   Ridley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P.,
RA   Wedler H., Wedler E., Wambutt R., Weitzenegger T., Pohl T.M.,
RA   Terryn N., Gielen J., Villarroel R., de Clerck R., van Montagu M.,
RA   Lecharny A., Auborg S., Gy I., Kreis M., Lao N., Kavanagh T.,
RA   Hempel S., Kotter P., Entian K.-D., Rieger M., Schaeffer M., Funk B.,
RA   Mueller-Auer S., Silvey M., James R., Montfort A., Pons A.,
RA   Puigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P.,
RA   Piravandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moores T.,
RA   Jones J.D.G., Eneva T., Palme K., Benes V., Rechman S., Ansorge W.,
RA   Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W.,
RA   Klosterman S., Schueller C., Chalwatzis N.;
RT   "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT   Arabidopsis thaliana.";
RL   Nature 391:485-488(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083488; PubMed=10617198;
RA   Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA   Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA   Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA   Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA   Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA   Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA   Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA   Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA   Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA   Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
```

```
RA   Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA   Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA   Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA   Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA   De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA   Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA   Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA   Pettett A., Rajandream M.-A., Lyne M., Benes V., Rechmann S.,
RA   Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA   Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA   Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA   Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA   Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA   Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA   Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA   Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA   Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA   Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA   Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA   Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA   Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA   Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA   Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA   Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA   Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA   Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA   Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA   Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA   Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA   Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA   Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA   Chen E., Marra M., Martienssen R., McCombie W.R.;
RT   "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:769-777(1999).
RN   [4]
RP   SEQUENCE OF 19-485 FROM N.A.
RC   STRAIN=CV. LANDSBERG ERECTA;
RA   Zhang H., Forde B.G.;
RT   "Identification of novel nitrate-inducible genes from Arabidopsis.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC       S-ADENOSYL-L-METHININE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC       THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC       CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC       CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC       ADENOSINE + L-HOMOCYSTEINE.
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z97335; CAB10173.1; -.
DR   EMBL; AF059581; AAC14714.1; -.
DR   EMBL; AL161537; CAB78436.1; -.
DR   EMBL; Z97059; CAB09795.1; -.
DR   InterPro; IPR000043; Ado_hcyase.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism.
FT   NP_BIND     263     294       NAD (POTENTIAL).
FT   CONFLICT     80      80       E -> Q (IN REF. 4).
FT   CONFLICT     96      96       A -> R (IN REF. 4).
FT   CONFLICT    392     392       E -> Q (IN REF. 4).
FT   CONFLICT    460     460       T -> R (IN REF. 4).
SQ   SEQUENCE   485 AA;  53378 MW;  1113270A0F46C86C CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     679 DATLLIHE 686
           ||||||||
Db     142 DATLLIHE 149


RESULT   5
SAHH_CATRO
ID   SAHH_CATRO     STANDARD;      PRT;   485 AA.
AC   P35007;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE
DE   HYDROLASE) (ADOHCYASE).
GN   SAHH.
OS   Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
OC   Vinceae; Catharanthus.
OX   NCBI_TaxID=4058;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94218405; PubMed=8165255;
RA   Schroeder G., Waitz A., Hotze M., Schroeder J.;
RT   "cDNA for S-adenosyl-L-homocysteine hydrolase from Catharanthus
RT   roseus.";
RL   Plant Physiol. 104:1099-1100(1994).
CC   -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC       S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC       THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC       CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC       CONCENTRATION OF ADENOSYLHOMOCYSTEINE.
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC       ADENOSINE + L-HOMOCYSTEINE.
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- INDUCTION: BY STRESS.
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z26881; CAA81527.1; -.
DR   PIR; S38379; S38379.
DR   InterPro; IPR000043; Ado_hcyase.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism.
FT   NP_BIND     263     294       NAD (POTENTIAL).
SQ   SEQUENCE   485 AA;  53233 MW;  50CC0E99A9F66C51 CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
```

```
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    679 DATLLIHE 686
          ||||||||
Db    142 DATLLIHE 149


RESULT   6
SAHH_MESCR
ID   SAHH_MESCR     STANDARD;      PRT;   485 AA.
AC   P93253;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE
DE   HYDROLASE) (ADOHCYASE).
GN   SAHH.
OS   Mesembryanthemum crystallinum (Common ice plant).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
OX   NCBI_TaxID=3544;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Michalowski C.B., Bohnert H.J.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC       S-ADENOSYL-L-METHININE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC       THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC       CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC       CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC       ADENOSINE + L-HOMOCYSTEINE.
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U79766; AAB38499.1; -.
DR   InterPro; IPR000043; Ado_hcyase.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism.
FT   NP_BIND     263    294       NAD (POTENTIAL).
SQ   SEQUENCE   485 AA;  53178 MW;  2C3B339BD4F7BAE6 CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    679 DATLLIHE 686
          ||||||||
Db    142 DATLLIHE 149


RESULT   7
SAHH_PETCR
ID   SAHH_PETCR     STANDARD;      PRT;   485 AA.
AC   Q01781;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE
DE   HYDROLASE) (ADOHCYASE).
GN   SAHH OR SHH.
OS   Petroselinum crispum (Parsley) (Petroselinum hortense).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.
OX   NCBI_TaxID=4043;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RX   MEDLINE=92262510; PubMed=1374911;
RA   Kawalleck P., Plesch G., Hahlbrock K., Somssich I.E.;
RT   "Induction by fungal elicitor of S-adenosyl-L-methionine synthetase
RT   and S-adenosyl-L-homocysteine hydrolase mRNAs in cultured cells and
RT   leaves of Petroselinum crispum.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:4713-4717(1992).
RN   [2]
RP   SEQUENCE OF 259-485 FROM N.A.
RA   Somssich I.E., Bollmann J., Hahlbrock K., Kombrink E., Schulz W.;
RT   "Differential early activation of defense-related genes in
RT   elicitor-treated parsley cells.";
RL   Plant Mol. Biol. 12:227-234(1989).
CC   -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC       S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC       THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC       CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC       CONCENTRATION OF ADENOSYLHOMOCYSTEINE.
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC       ADENOSINE + L-HOMOCYSTEINE.
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- TISSUE SPECIFICITY: MAINLY IN FLORAL BUDS AND STEMS.
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81885; AAA33856.1; -.
DR   EMBL; M62756; AAA33855.1; -.
DR   InterPro; IPR000043; Ado_hcyase.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism.
FT   NP_BIND     263    294       NAD (POTENTIAL).
FT   CONFLICT    346    346       C -> D (IN REF. 1; AAA33855).
FT   CONFLICT    439    439       L -> C (IN REF. 1; AAA33855).
SQ   SEQUENCE   485 AA;  53181 MW;  05E926516C2E08E6 CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    679 DATLLIHE 686
          ||||||||
Db    142 DATLLIHE 149


RESULT   8
SAHH_PHASS
ID   SAHH_PHASS     STANDARD;      PRT;   485 AA.
AC   P50249;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE
```

```
DE   HYDROLASE) (ADOHCYASE).
GN   SAHH.
OS   Phalaenopsis sp.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Orchidaceae;
OC   Epidendroideae; higher Epidendroideae; Vandeae; Aeridinae;
OC   Phalaenopsis.
OX   NCBI_TaxID=36900;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95177653; PubMed=7872785;
RA   Preisig-Mueller R., Gnau P., Kindl H.;
RT   "The inducible 9, 10-dihydrophenanthrene pathway: characterization
RT   and expression of bibenzyl synthase and S-adenosylhomocysteine
RT   hydrolase.";
RL   Arch. Biochem. Biophys. 317:201-207(1995).
CC   -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC       S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC       THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC       CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC       CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC       ADENOSINE + L-HOMOCYSTEINE.
CC   -!- COFACTOR: NAD (BY SIMILARITY).
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- INDUCTION: BY INFECTION WITH B.CINEREA.
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X79905; CAA56278.1; -.
DR   InterPro; IPR000043; Ado_hcyase.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism.
FT   NP_BIND     263    294       NAD (POTENTIAL).
SQ   SEQUENCE   485 AA;  53141 MW;  1EB2CA5AD63AF233 CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    679 DATLLIHE 686
          ||||||||
Db    142 DATLLIHE 149


RESULT   9
SAHH_TOBAC
ID   SAHH_TOBAC     STANDARD;      PRT;   485 AA.
AC   P50248;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE
DE   HYDROLASE) (ADOHCYASE) (CYTOKININ BINDING PROTEIN CBP57).
GN   SAHH.
OS   Nicotiana tabacum (Common tobacco), and
OS   Nicotiana sylvestris (Wood tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097, 4096;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=N.tabacum; STRAIN=CV. BRIGHT YELLOW 4;
RA   Tanaka H., Masuta C., Kataoka J., Kuwata S., Koiwai A., Noma M.;
RT   "Inducible expression by plant hormones of S-adenosyl-homocysteine
RT   hydrolase gene from Nicotiana tabacum during early flower bud
RT   formation in vitro.";
RL   Plant Sci. 113:167-174(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=N.sylvestris;
RA   Mitsui S., Wakasugi T., Sugiura M.;
RT   "A cDNA encoding the 57 kDa subunit of a cytokinin-binding protein
RT   complex from tobacco: the subunit has high homology to S-adenosyl-L-
RT   homocystein hydrolase.";
RL   Plant Cell Physiol. 34:1089-1096(1993).
CC   -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC       S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC       THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC       CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC       CONCENTRATION OF ADENOSYLHOMOCYSTEINE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC       ADENOSINE + L-HOMOCYSTEINE.
CC   -!- COFACTOR: NAD (BY SIMILARITY).
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D45204; BAA08142.1; -.
DR   EMBL; D49804; BAA23164.1; -.
DR   EMBL; D16138; BAA03709.1; -.
DR   InterPro; IPR000043; Ado_hcyase.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism.
FT   NP_BIND     263    294       NAD (POTENTIAL).
SQ   SEQUENCE   485 AA;  53104 MW;  AA6D6844E9DF0A5C CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    679 DATLLIHE 686
          ||||||||
Db    142 DATLLIHE 149


RESULT  10
SAHH_WHEAT
ID   SAHH_WHEAT     STANDARD;      PRT;   485 AA.
AC   P32112;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   ADENOSYLHOMOCYSTEINASE (EC 3.3.1.1) (S-ADENOSYL-L-HOMOCYSTEINE
DE   HYDROLASE) (ADOHCYASE).
GN   SAHH OR SHH OR SH6.2.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Richards K.D., Gardner R.C.;
RL   Submitted (XXX-1993) to the EMBL/GenBank/DDBJ databases.
```

```
CC   -!- FUNCTION: ADENOSYLHOMOCYSTEINE IS A COMPETITIVE INHIBITOR OF
CC       S-ADENOSYL-L-METHIONINE-DEPENDENT METHYL TRANSFERASE REACTIONS;
CC       THEREFORE ADENOSYLHOMOCYSTEINASE MAY PLAY A KEY ROLE IN THE
CC       CONTROL OF METHYLATIONS VIA REGULATION OF THE INTRACELLULAR
CC       CONCENTRATION OF ADENOSYLHOMOCYSTEINE.
CC   -!- CATALYTIC ACTIVITY: S-ADENOSYL-L-HOMOCYSTEINE + H(2)O =
CC       ADENOSINE + L-HOMOCYSTEINE.
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L11872; AAA34303.1; -.
DR   InterPro; IPR000043; Ado_hcyase.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism.
FT   NP_BIND     263    294       NAD (POTENTIAL).
SQ   SEQUENCE   485 AA;  53436 MW;  97FE208E3A6874ED CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 485;
  Best Local Similarity   100.0%;  Pred. No. 8.2;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    679 DATLLIHE 686
          ||||||||
Db    142 DATLLIHE 149


RESULT  11
Y167_HUMAN
ID   Y167_HUMAN     STANDARD;      PRT;   836 AA.
AC   Q99490; O00578;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL PROTEIN KIAA0167.
GN   KIAA0167.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97336055; PubMed=9192850;
RA   Elkahloun A.G., Krizman D.B., Wang Z., Hofmann T.A., Roe B.,
RA   Meltzer P.S.;
RT   "Transcript mapping in a 46-kb sequenced region at the core of 12q13.3
RT   amplification in human cancers.";
RL   Genomics 42:295-301(1997).
CC   -!- SIMILARITY: CONTAINS 2 ANK REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D79989; BAA11484.1; -.
DR   EMBL; U81031; AAC39522.1; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR003575; Small_GTPase.
DR   InterPro; IPR001164; Znf_GCS.
DR   Pfam; PF00023; ank; 2.
DR   Pfam; PF01412; ArfGap; 1.
DR   Pfam; PF00169; PH; 1.
DR   PRINTS; PR00405; REVINTRACTNG.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   SMART; SM00248; ANK; 1.
DR   SMART; SM00105; ArfGap; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00010; small_GTPase; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 1.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
KW   Hypothetical protein; Repeat; ANK repeat.
FT   DOMAIN      340    554       PH.
FT   REPEAT      734    763       ANK 1.
FT   REPEAT      767    796       ANK 2.
FT   CONFLICT     53     53       MISSING (IN REF. 2).
FT   CONFLICT    432    433       MISSING (IN REF. 2).
SQ   SEQUENCE   836 AA;  90536 MW;  B8E42631BC06D5DC CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 836;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    333 ADAPVALV 340
          ||||||||
Db    829 ADAPVALV 836


RESULT  12
MDR2_CRIGR
ID   MDR2_CRIGR     STANDARD;      PRT;  1276 AA.
AC   P21449;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MULTIDRUG RESISTANCE PROTEIN 2 (P-GLYCOPROTEIN 2).
GN   PGY2 OR PGP2.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92135896; PubMed=1685679;
RA   Endicott J.A., Sarangi F., Ling V.;
RT   "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
RT   gene family.";
RL   DNA Seq. 2:89-101(1991).
RN   [2]
RP   SEQUENCE OF 622-1276 FROM N.A.
RX   MEDLINE=88122132; PubMed=2893255;
RA   Endicott J.A., Juranka P.F., Sarangi F., Gerlach J.H., Deuchars K.L.,
RA   Ling V.;
RT   "Simultaneous expression of two P-glycoprotein genes in
RT   drug-sensitive Chinese hamster ovary cells.";
RL   Mol. Cell. Biol. 7:4075-4081(1987).
```

```
CC   -!- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC       DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
CC       CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
CC       CANNOT.
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS). MDR SUBFAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M60041; AAA68884.1; -.
DR   EMBL; M17896; AAA37007.1; -.
DR   PIR; B27126; DVHY2C.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001140; ABC_trnsportr_tmem.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   Pfam; PF00005; ABC_tran; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW   Multigene family.
FT   DOMAIN        1     51       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     52     72       POTENTIAL.
FT   TRANSMEM    119    139       POTENTIAL.
FT   TRANSMEM    188    208       POTENTIAL.
FT   TRANSMEM    215    235       POTENTIAL.
FT   TRANSMEM    296    316       POTENTIAL.
FT   TRANSMEM    325    345       POTENTIAL.
FT   DOMAIN      346    708       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    709    729       POTENTIAL.
FT   TRANSMEM    755    775       POTENTIAL.
FT   TRANSMEM    831    851       POTENTIAL.
FT   TRANSMEM    852    872       POTENTIAL.
FT   TRANSMEM    935    955       POTENTIAL.
FT   TRANSMEM    972    992       POTENTIAL.
FT   DOMAIN      993   1276       CYTOPLASMIC (POTENTIAL).
FT   NP_BIND     426    433       ATP (POTENTIAL).
FT   NP_BIND    1068   1075       ATP (POTENTIAL).
FT   REPEAT        1    635
FT   REPEAT      636   1276
SQ   SEQUENCE   1276 AA;  141057 MW;  5096B1385628812D CRC64;


  Query Match             1.0%;  Score 8;  DB 1;  Length 1276;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    535 VLGTLAAV 542
          ||||||||
Db     52 VLGTLAAV 59


RESULT  13
YCX4_MESVI
ID   YCX4_MESVI     STANDARD;      PRT;    71 AA.
AC   Q9MUR2;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   HYPOTHETICAL 8.1 KDA PROTEIN IN NDHF-PSBD INTERGENIC REGION.
OS   Mesostigma viride.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Mesostigmatophyceae;
OC   Mesostigmatales; Mesostigmataceae; Mesostigma.
OX   NCBI_TaxID=41882;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIES-296;
RX   MEDLINE=20150907; PubMed=10688199;
RA   Lemieux C., Otis C., Turmel M.;
RT   "Ancestral chloroplast genome in Mesostigma viride reveals an early
RT   branch of green plant evolution.";
RL   Nature 403:649-652(2000).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF166114; AAF43891.1; -.
KW   Hypothetical protein; Chloroplast.
SQ   SEQUENCE   71 AA;  8147 MW;  9608379E175B0319 CRC64;


  Query Match             0.8%;  Score 7;  DB 1;  Length 71;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    623 RLISSLL 629
          |||||||
Db     10 RLISSLL 16


RESULT  14
IR08_HCMVA
ID   IR08_HCMVA     STANDARD;      PRT;   129 AA.
AC   P16806;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   HYPOTHETICAL PROTEIN IRL8 (TRL8).
OS   Human cytomegalovirus (strain AD169).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=10360;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90269039; PubMed=2161319;
RA   Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA   Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,
RA   Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
RT   "Analysis of the protein-coding content of the sequence of human
RT   cytomegalovirus strain AD169.";
RL   Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X17403; CAA35456.1; -.
DR   EMBL; X17403; CAA35302.1; -.
DR   PIR; S09757; S09757.
KW   Hypothetical protein.
SQ   SEQUENCE   129 AA;  14301 MW;  EDDEEF121B86F0D3 CRC64;


  Query Match             0.8%;  Score 7;  DB 1;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy    779 EERREKR 785
          |||||||
Db     12 EERREKR 18


RESULT  15
VG42_HSVI1
ID   VG42_HSVI1     STANDARD;      PRT;   129 AA.
AC   Q00101;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-DEC-1992 (Rel. 24, Last annotation update)
DE   HYPOTHETICAL GENE 42 PROTEIN.
GN   42.
OS   Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   unclassified Herpesviridae.
OX   NCBI_TaxID=10401;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AUBURN 1;
RX   MEDLINE=92087490; PubMed=1727613;
RA   Davison A.J.;
RT   "Channel catfish virus: a new type of herpesvirus.";
RL   Virology 186:9-14(1992).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M75136; AAA88145.1; -.
DR   PIR; G36790; G36790.
KW   Hypothetical protein.
SQ   SEQUENCE   129 AA;  14346 MW;  DA7ECE85FECD4E9E CRC64;


  Query Match             0.8%;  Score 7;  DB 1;  Length 129;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    230 QRRGVRD 236
          |||||||
Db     32 QRRGVRD 38


Search completed: March 29, 2002, 10:24:41
Job time: 138 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 29, 2002, 10:21:23 ; Search time 21.56 Seconds
(without alignments)
2918.376 Million cell updates/sec

Title: US-09-988-687-2
Perfect score: 826
Sequence: 1 MWALCSLLRSAAGRTMSQGR.....EPQQKRAHTEEPQAKKVRAQ 826

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	11	1.3	809	2	T38051	hypothetical prote
2	10	1.2	837	2	F96561	unknown protein [i
3	9	1.1	838	2	S38156	hypothetical prote
4	8	1.0	147	2	D75073	hit-like protein P
5	8	1.0	216	2	T23258	hypothetical prote
6	8	1.0	255	2	T35217	hypothetical prote
7	8	1.0	260	2	G75111	hypothetical prote
8	8	1.0	280	1	A71022	hypothetical prote
9	8	1.0	307	2	A71057	probable sulfatase
10	8	1.0	440	2	C85785	probable thiosulfa
11	8	1.0	445	1	DCUTOB	ornithine decarbox
12	8	1.0	485	2	C71400	adenosylhomocystei
13	8	1.0	485	2	S71621	adenosylhomocystei
14	8	1.0	485	2	S38379	adenosylhomocystei
15	8	1.0	485	2	T06764	adenosylhomocystei
16	8	1.0	643	2	F84117	hypothetical prote
17	8	1.0	806	2	T35640	probable sensor ki
18	8	1.0	989	2	C83035	hypothetical prote
19	8	1.0	1630	2	T00390	KIAA0614 protein -
20	7	0.8	106	2	T36932	probable integral
21	7	0.8	122	2	S40351	Ig kappa chain V-J
22	7	0.8	129	2	S09757	hypothetical prote
23	7	0.8	129	2	G36790	hypothetical prote
24	7	0.8	139	2	C83526	hypothetical prote
25	7	0.8	144	2	B70983	hypothetical prote
26	7	0.8	144	2	T33858	hypothetical prote
27	7	0.8	144	2	H82837	conserved hypothet
28	7	0.8	148	2	B72513	hypothetical prote
29	7	0.8	149	2	C69393	transcription regu
30	7	0.8	156	2	T12893	hypothetical prote
31	7	0.8	167	2	T44271	glutathione peroxi
32	7	0.8	167	2	T35112	probable transmemb
33	7	0.8	171	2	A45562	merozoite surface
34	7	0.8	189	2	T48828	hypothetical prote
35	7	0.8	197	2	B49247	merozoite surface
36	7	0.8	197	2	C49247	merozoite surface
37	7	0.8	209	2	E82388	conserved hypothet
38	7	0.8	211	2	T10392	late expression fa
39	7	0.8	216	2	C64994	InaA protein - Esc
40	7	0.8	216	2	B85863	hypothetical prote
41	7	0.8	239	2	T01599	hypothetical prote
42	7	0.8	241	2	H31844	KorA protein - Str
43	7	0.8	247	2	D84304	hypothetical prote
44	7	0.8	250	2	A83310	hypothetical prote
45	7	0.8	250	2	T36131	hypothetical prote

ALIGNMENTS

RESULT 1
T38051
hypothetical protein SPAC1D4.10 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38051
R;Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1995
A;Reference number: Z21765
A;Accession: T38051
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-809 <LYE>
A;Cross-references: EMBL:Z69239; PIDN:CAA93219.1; GSPDB:GN00066; SPDB:SPAC1D4.10
A;Experimental source: strain 972h-; cosmid c1D4
C;Genetics:
A;Gene: SPDB:SPAC1D4.10
A;Map position: 1
A;Introns: 129/1

Query Match 1.3%; Score 11; DB 2; Length 809;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 681 TLLIHEATLED 691
||||||||||||
Db 704 TLLIHEATLED 714

RESULT 2
F96561
unknown protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: F96561
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallc
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: F96561
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-837 <STO>
A;Cross-references: GB:AE005173; NID:g6850339; PIDN:AAF29402.1; GSPDB:GN00141

```
C;Genetics:
A;Gene: F9I5.1
A;Map position: 1


  Query Match             1.2%;  Score 10;  DB 2;  Length 837;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    511 SLLLDCGEGT 520
          ||||||||||
Db    515 SLLLDCGEGT 524


RESULT   3
S38156
hypothetical protein YKR079c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Oct-1999
C;Accession: S38156
R;Pohl, T.M.; Pohl, F.M.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37897
A;Accession: S38156
A;Molecule type: DNA
A;Residues: 1-838 <POH>
A;Cross-references: EMBL:Z28304; NID:g486556; PID:g486557; GSPDB:GN00011; MIPS:YKR079c
A;Experimental source: strain S288C
C;Genetics:
A;Gene: MIPS:YKR079c.
A;Map position: 11R


  Query Match             1.1%;  Score 9;  DB 2;  Length 838;
  Best Local Similarity   100.0%;  Pred. No. 2.1;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    682 LLIHEATLE 690
          |||||||||
Db    717 LLIHEATLE 725


RESULT   4
D75073
hit-like protein PAB0657 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: D75073
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: D75073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-147 <KAW>
A;Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB49889.1; PID:g545840
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0657
C;Superfamily: protein kinase C inhibitor; histidine triad homology
F;3-99/Domain: histidine triad homology <HIT>


  Query Match             1.0%;  Score 8;  DB 2;  Length 147;
  Best Local Similarity   100.0%;  Pred. No. 5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    158 LKGIELAV 165
          ||||||||
Db     57 LKGIELAV 64


RESULT   5
T23258
hypothetical protein K03A11.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T23258
R;Swinburne, J.
submitted to the EMBL Data Library, July 1996
A;Reference number: Z19717
A;Accession: T23258
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-216 <WIL>
A;Cross-references: EMBL:Z77133; PIDN:CAB00867.1; GSPDB:GN00028; CESP:K03A11.3
A;Experimental source: clone K03A11
C;Genetics:
A;Gene: CESP:K03A11.3
A;Map position: X
A;Introns: 136/3; 168/3


  Query Match             1.0%;  Score 8;  DB 2;  Length 216;
  Best Local Similarity   100.0%;  Pred. No. 7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     78 SEFNRYLF 85
          ||||||||
Db     75 SEFNRYLF 82


RESULT   6
T35217
hypothetical protein SC5C7.10c SC5C7.10c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C;Accession: T35217
R;Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21572
A;Accession: T35217
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-255 <SEE>
A;Cross-references: EMBL:AL031515; PIDN:CAA20622.1; GSPDB:GN00070; SCOEDB:SC5C7.10c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC5C7.10c
C;Superfamily: conserved hypothetical protein MJ1163


  Query Match             1.0%;  Score 8;  DB 2;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 8.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    544 VSHLHADH 551
          ||||||||
Db     53 VSHLHADH 60


RESULT   7
G75111
hypothetical protein PAB1906 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: G75111
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s
A;Reference number: A75001
A;Accession: G75111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-260 <KAW>
```

A;Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49608.1; PID:g545811
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB1906
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0209

Query Match 1.0%; Score 8; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 671 EALVRMGK 678
||||||||
Db 167 EALVRMGK 174

RESULT 8
A71022
hypothetical protein PH1470 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C;Accession: A71022
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: A71022
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-280 <KAW>
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30577.1; PID:g3257894
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1470
C;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0209

Query Match 1.0%; Score 8; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 671 EALVRMGK 678
||||||||
Db 188 EALVRMGK 195

RESULT 9
A71057
probable sulfatase - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: A71057
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137
A;Accession: A71057
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-307 <KAW>
A;Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30251.1; PID:g3257568
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1151
C;Superfamily: conserved hypothetical protein MJ1502

Query Match 1.0%; Score 8; DB 2; Length 307;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 660 KVVYSGDT 667
||||||||
Db 202 KVVYSGDT 209

RESULT 10
C85785
probable thiosulfate sulfur transferase Z2789 [imported] - Escherichia coli (strain O
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C;Accession: C85785
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85785
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-440 <STO>
A;Cross-references: GB:AE005174; NID:g12515779; PIDN:AAG56743.1; GSPDB:GN00145; UWGP:
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z2789

Query Match 1.0%; Score 8; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 332 KADAPVAL 339
||||||||
Db 99 KADAPVAL 106

RESULT 11
DCUTOB
ornithine decarboxylase (EC 4.1.1.17) - Trypanosoma brucei
C;Species: Trypanosoma brucei
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
C;Accession: A29309
R;Phillips, M.A.; Coffino, P.; Wang, C.C.
J. Biol. Chem. 262, 8721-8727, 1987
A;Title: Cloning and sequencing of the ornithine decarboxylase gene from Trypanosoma
A;Reference number: A29309; MUID:87250494
A;Accession: A29309
A;Molecule type: DNA
A;Residues: 1-445 <PHI>
A;Cross-references: GB:J02771; NID:g162173; PIDN:AAA30218.1; PID:g162174
C;Superfamily: ornithine decarboxylase
C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; polyamine biosynthesi
F;89/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted
F;380/Active site: Cys (shared with dimeric partner) #status predicted

Query Match 1.0%; Score 8; DB 1; Length 445;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 534 RVLGTLAA 541
||||||||
Db 95 RVLGTLAA 102

RESULT 12
C71400
adenosylhomocysteinase (EC 3.3.1.1) [similarity] - Arabidopsis thaliana
N;Alternate names: protein DL3010W; S-adenosyl-L-homocysteine hydrolase
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 02-Sep-2000

C;Accession: C71400; T51608
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
C.; Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A;Reference number: A71400; MUID:98121113
A;Accession: C71400
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-485 <BEV>
A;Cross-references: GB:Z97335; NID:g2244747; PIDN:CAB10173.1; PID:g2244750
R;Belbahri, L.; Elleuch, H.; Villaroel, R.; Inze, D.; Thomas, D.; Thomasset, B.
Plant Physiol. 121, 313, 1999
A;Title: The Isolation of an Arabidopsis thaliana cDNA clone encoding S-adenosyl-L-homoc
A;Reference number: Z25409
A;Accession: T51608
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-485 <BEL>
A;Cross-references: EMBL:AF059581; PIDN:AAC14714.1
A;Experimental source: cultivar Landsberg erecta
C;Genetics:
A;Gene: SAHH
A;Map position: 4COP9-4G3845
A;Note: DL3010W
C;Superfamily: adenosylhomocysteinase
C;Keywords: NAD; thioether hydrolase
F;263-292/Region: beta-alpha-beta NAD nucleotide-binding fold
F;86,120/Active site: Cys #status predicted

Query Match 1.0%; Score 8; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 679 DATLLIHE 686
||||||||
Db 142 DATLLIHE 149

RESULT 13
S71621
adenosylhomocysteinase (EC 3.3.1.1) - Phalaenopsis sp.
N;Alternate names: S-adenosyl-L-homocysteine hydrolase
C;Species: Phalaenopsis sp.
C;Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 22-Jun-1999
C;Accession: S71621
R;Preisig-Mueller, R.; Gnau, P.; Kindl, H.
Arch. Biochem. Biophys. 317, 201-207, 1995
A;Title: The inducible 9,10-dihydrophenanthrene pathway: characterization and expression
A;Reference number: S71619; MUID:95177653
A;Accession: S71621
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-485 <PRE>
A;Cross-references: EMBL:X79905; NID:g758246; PIDN:CAA56278.1; PID:g758247
A;Experimental source: clone SHH511
C;Superfamily: adenosylhomocysteinase
C;Keywords: NAD; thioether hydrolase
F;263-292/Region: beta-alpha-beta NAD nucleotide-binding fold
F;86,120/Active site: Cys #status predicted

Query Match 1.0%; Score 8; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 679 DATLLIHE 686
||||||||
Db 142 DATLLIHE 149

RESULT 14
S38379
adenosylhomocysteinase (EC 3.3.1.1) - Madagascar periwinkle
N;Alternate names: S-adenosyl-L-homocysteine hydrolase
C;Species: Catharanthus roseus (Madagascar periwinkle)
C;Date: 22-Jan-1994 #sequence_revision 10-Nov-1995 #text_change 23-Jul-1999
C;Accession: S38379
R;Schroeder, G.; Waitz, A.; Hotze, M.; Schroeder, J.
submitted to the EMBL Data Library, October 1993
A;Description: cDNA for stress-induced S-adenosyl-L-homocysteine hydrolase from Madag
A;Reference number: S38379
A;Accession: S38379
A;Molecule type: DNA
A;Residues: 1-485 <SCH>
A;Cross-references: EMBL:Z26881; NID:g407411; PIDN:CAA81527.1; PID:g407412
C;Genetics:
A;Gene: SAHH
C;Superfamily: adenosylhomocysteinase
C;Keywords: NAD; thioether hydrolase
F;263-292/Region: beta-alpha-beta NAD nucleotide-binding fold
F;86,120/Active site: Cys #status predicted

Query Match 1.0%; Score 8; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 679 DATLLIHE 686
||||||||
Db 142 DATLLIHE 149

RESULT 15
T06764
adenosylhomocysteinase (EC 3.3.1.1) - wheat
N;Alternate names: S-adenosyl-L-homocysteine hydrolase
C;Species: Triticum aestivum (common wheat)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C;Accession: T06764
R;Richards, K.D.; Gardner, R.C.
submitted to the EMBL Data Library, March 1993
A;Description: The influence of aluminium on histone, heat shock and S-adenosyl-L-hom
A;Reference number: Z15795
A;Accession: T06764
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-485 <RIC>
A;Cross-references: EMBL:L11872; NID:g170772; PIDN:AAA34303.1; PID:g170773
C;Genetics:
A;Gene: SH6.2
C;Superfamily: adenosylhomocysteinase
C;Keywords: NAD; thioether hydrolase
F;263-292/Region: beta-alpha-beta NAD nucleotide-binding fold
F;86,120/Active site: Cys #status predicted

Query Match 1.0%; Score 8; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 679 DATLLIHE 686
||||||||
Db 142 DATLLIHE 149

Search completed: March 29, 2002, 10:23:41
Job time: 138 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 29, 2002, 10:20:33 ; Search time 28.96 Seconds
(without alignments)
2112.726 Million cell updates/sec

Title: US-09-988-687-2
Perfect score: 826
Sequence: 1 MWALCSLLRSAAGRTMSQGR.....EPQQKRAHTEEPQAKKVRAQ 826

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_1101:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	826	100.0	826	21	AAB07228	Human prostate can
2	624	75.5	826	21	AAY99850	Human sulphatase G
3	505	61.1	606	21	AAB42236	Human ORFX ORF2000
4	430	52.1	854	22	AAB92918	Human protein sequ
5	15	1.8	73	21	AAB07230	Murine prostate ca
6	10	1.2	118	21	AAG18567	Arabidopsis thalia
7	10	1.2	154	21	AAG18566	Arabidopsis thalia
8	9	1.1	838	22	AAB49964	S. cerevisiae YKR0
9	8	1.0	141	21	AAB25679	Human secreted pro
10	8	1.0	243	21	AAB43610	Human cancer assoc
11	8	1.0	245	22	AAG75627	Human colon cancer
12	8	1.0	377	21	AAB42637	Human ORFX ORF2401
13	8	1.0	385	22	AAB85355	Human phosphatase
14	8	1.0	394	21	AAG08256	Arabidopsis thalia
15	8	1.0	422	21	AAG08255	Arabidopsis thalia
16	8	1.0	449	21	AAG22958	Arabidopsis thalia
17	8	1.0	450	21	AAG34061	Zea mays protein f
18	8	1.0	467	21	AAG22957	Arabidopsis thalia
19	8	1.0	467	21	AAG40087	Arabidopsis thalia
20	8	1.0	467	21	AAG41408	Arabidopsis thalia
21	8	1.0	485	17	AAW01459	Arabidopsis S-aden
22	8	1.0	485	21	AAG34060	Zea mays protein f
23	8	1.0	485	21	AAG40086	Arabidopsis thalia
24	8	1.0	485	21	AAG41407	Arabidopsis thalia
25	8	1.0	486	13	AAR26500	Prod. of Nicotiana
26	8	1.0	489	21	AAG22956	Arabidopsis thalia
27	8	1.0	497	21	AAG41406	Arabidopsis thalia
28	8	1.0	508	21	AAG40085	Arabidopsis thalia
29	8	1.0	509	21	AAG08254	Arabidopsis thalia
30	8	1.0	510	21	AAG34059	Zea mays protein f
31	8	1.0	790	21	AAG49847	Arabidopsis thalia
32	8	1.0	848	21	AAG49846	Arabidopsis thalia
33	8	1.0	922	21	AAG49845	Arabidopsis thalia
34	8	1.0	1632	22	AAM39023	Human polypeptide
35	7	0.8	20	22	AAM17295	Peptide #3729 enco
36	7	0.8	20	22	AAM29793	Peptide #3830 enco
37	7	0.8	52	21	AAB27820	Human secreted pro
38	7	0.8	57	22	AAM19543	Peptide #5977 enco
39	7	0.8	57	22	AAM32583	Peptide #6620 enco
40	7	0.8	81	21	AAB33060	Pinus radiata tran
41	7	0.8	86	21	AAY87065	Human secreted pro
42	7	0.8	86	22	AAE06042	Human gene 2 encod
43	7	0.8	99	21	AAB53940	Human colon cancer
44	7	0.8	108	21	AAY87157	Human secreted pro
45	7	0.8	108	22	AAE06134	Human gene 2 encod

ALIGNMENTS

RESULT 1
AAB07228
ID AAB07228 standard; protein; 826 AA.
XX
AC AAB07228;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human prostate cancer predisposing protein HPC2.
XX
KW Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW gene therapy; peptide therapy; drug design.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
DR N-PSDB; AAA58453.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Claim 2; Page 96-98; 157pp; English.

```
XX
CC   The present sequence is the protein produced from the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This protein and its gene can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs.
XX
SQ   Sequence   826 AA;


  Query Match            100.0%;  Score 826;  DB 21;  Length 826;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  826;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mwalcsllrsaagrtmsqgrtisqaparrerprkdplrhlrtrekrgpsgcsggpntvyl 60

Qy      61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 qvvaagsrdsgaalyvfsefnrylfncgegvqrlmqehklkvarldnifltrmhwsnvgg 120

Qy     121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 lsgmiltlketglpkcvlsgppqlekyleaikifsgplkgielavrphsapeyedetmtv 180

Qy     181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 yqipihseqrrgkhqpwqsperplsrlsperssdsesnenephlphgvsqrrgvrdsslv 240

Qy     241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 vaficklhlkrgnflvlkakemglpvgtaaiapiiaavkdgksithegreilaeelctpp 300

Qy     301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 dpgaafvvvecpdesfiqpicenatfqryqgkadapvalvvhmapasvlvdsryqqwmer 360

Qy     361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 fgpdtqhlvlnencasvhnlrshkiqtqlnlihpdifplltsfrckkegptlsvpmvqge 420

Qy     421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 cllkyqlrprrewqrdaiitcnpeefivealqlpnfqqsvqeyrrsaqdgpapaekrsqy 480

Qy     481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 peiiflgtgsaipmkirnvsatlvnispdtslllcgegtfgqlcrhygdqvdrvlgtla 540

Qy     541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 avfvshlhadhhtglpsillqreralaslgkplhpllvvapnqlkawlqqyhnqcqevlh 600

Qy     601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 hismipakclqegaeisspaverlissllrtcdleefqtclvrhckhafgcalvhtsgwk 660

Qy     661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 vvysgdtmpcealvrmgkdatlliheatledgleeeavekthsttsqaisvgmrmnaefi 720

Qy     721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEE 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 mlnhfsqryakvplfspnfsekvgvafdhmkvcfgdfptmpklippl kalfagdieemee 780

Qy     781 RREKRELRQVRAALLSRELAGGLEDGEPQQKRAHTEEPQAKKVRAQ 826
           |||||||||||||||||||||||||||||||||||||||||||||
Db     781 rrekrelrqvraallsrelaggledgepqqkrahteepqakkvraq 826
```

```
RESULT   2
AAY99850
ID   AAY99850 standard; Protein; 826 AA.
XX
AC   AAY99850;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Human sulphatase G.
XX
KW   Human sulphatase G; hSG; chromosome 17p11.2; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200034327-A1.
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999;   99WO-AU01092.
XX
PR   09-DEC-1998;   98AU-0007624.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Hopwood JJ,  Litjens T,  Hu RL;
XX
DR   WPI; 2000-431273/37.
DR   N-PSDB; AAA52810.
XX
PT   Novel isolated DNA sequence which encodes human sulfatase G or its
PT   fragment useful in gene therapy for treating patients suffering from
PT   sulfatase deficiency -
XX
PS   Claim 13; Fig 1; 33pp; English.
XX
CC   The present sequence is human sulphatase G (hSG). hSG is not a member of
CC   the well-characterised CTPSR sulphatase family. It belongs to a family
CC   showing sequence similarity to a sulphatase from the marine bacterium
CC   Alteromonas carrageenovora. The hSG gene contains 23 exons and is
CC   located at chromosome 17p11.2. The nucleotide sequence encoding
CC   this protein was identified in clone lambda29.1 of a human testes cDNA
CC   library. The clone was isolated using human EST sequences with sequence
CC   similarity to the non-CTPSR family as a probe to screen the library. The
CC   cDNA insert was subcloned and the DNA sequence of both strands was
CC   determined. hSG nucleotide and polypeptide sequences may be used to treat
CC   a patient suffering from hSG deficiency by replacing, repairing, or
CC   compensating for a DNA sequence within that patient's genome.
XX
SQ   Sequence   826 AA;


  Query Match            75.5%;  Score 624;  DB 21;  Length 826;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches  824;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mwalcsllrsaagrtmsqgrtisqaparrerprkdplrhlrtrekrgpsgcsggpntvyl 60

Qy      61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 qvvaagsrdsgaalyvfsefnrylfncgegvqrlmqehklkvarldnifltrmhwsnvgg 120

Qy     121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 lsgmiltlketglpkcvlsgppqlekyleaikifsgplkgielavrphsapeyedetmtv 180

Qy     181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     181 yqipihseqrrgkhqpwqsperplsrlsperssdselnenephlphgvsqrrgvrdsslv 240

Qy     241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    241 vaficklhlkrgnflvlkakemglpvgtaaiapiiaavkdgksithegreilaeelctpp 300

Qy    301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 dpgaafvvvecpdesfiqpicenatfqryqgkadapvalvvhmapasvlvdsryqqwmer 360

Qy    361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 fgpdtqhlvlnencasvhnlrshkiqtqlnlihpdifplltsfrckkegptlsvpmvqge 420

Qy    421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cllkyqlrprrewqrdaiitcnpeefivealqlpnfqqsvqeyrrsaqdgpapaekrsqy 480

Qy    481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 peiiflgtgsaipmkirnvsatlvnispdtslldcgegtfgqlcrhygdqvdrvlgtla 540

Qy    541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 tvfvshlhadhhtglpsillqreralaslgkplhpllvvapnqlkawlqqyhnqcqevlh 600

Qy    601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 hismipakclqegaeisspaverlissllrtcdleefqtclvrhckhafgcalvhtsgwk 660

Qy    661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 vvysgdtmpcealvrmgkdatlliheatledgleeeavekthsttsqaisvgmrmnaefi 720

Qy    721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEE 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 mlnhfsqryakvplfspnfsekvgvafdhmkvcfgdfptmpklipplkalfagdieemee 780

Qy    781 RREKRELRQVRAALLSRELAGGLEDGEPQQKRAHTEEPQAKKVRAQ 826
          |||||||||||||||||||||||||||||||||||||||||||||
Db    781 rrekrelrqvraallsrelaggledgepqqkrahteepqakkvraq 826


RESULT 3
AAB42236
ID   AAB42236 standard; Protein; 606 AA.
XX
AC   AAB42236;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF2000 polypeptide sequence SEQ ID NO:4000.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC76445.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 3181-3182; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   606 AA;


  Query Match             61.1%;  Score 505;  DB 21;  Length 606;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  505;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    262 MGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPIC 321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mglpvgtaaiapiiaavkdgksithegreilaeelctppdpgaafvvvecpdesfiqpic 60

Qy    322 ENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLR 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 enatfqryqgkadapvalvvhmapasvlvdsryqqwmerfgpdtqhlvlnencasvhnlr 120

Qy    382 SHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITC 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 shkiqtqlnlihpdifplltsfrckkegptlsvpmvqgecllkyqlrprrewqrdaiitc 180

Qy    442 NPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSA 501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 npeefivealqlpnfqqsvqeyrrsaqdgpapaekrsqypeiiflgtgsaipmkirnvsa 240

Qy    502 TLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQ 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tlvnispdtslllcgegtfgqlcrhygdqvdrvlgtlaavfvshlhadhhtglpsillq 300

Qy    562 RERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAV 621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 reralaslgkplhpllvvapnqlkawlqqyhnqcqevlhhismipakclqegaeisspav 360

Qy    622 ERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDAT 681
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 erlissllrtcdleefqtclvrhckhafgcalvhtsgwkvvysgdtmpcealvrmgkdat 420
```

```
Qy    682 LLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFSE 741
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 lliheatledgleeeavekthsttsqaisvgmrmnaefimlnhfsqryakvplfspnfse 480

Qy    742 KVGVAFDHMKVCFGDFPTMPKLIPP 766
          |||||||||||||||||||||||||
Db    481 kvgvafdhmkvcfgdfptmpklipp 505


RESULT   4
AAB92918
ID   AAB92918 standard; Protein; 854 AA.
XX
AC   AAB92918;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:11558.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11558; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   854 AA;

  Query Match             52.1%;  Score 430;  DB 22;  Length 854;
  Best Local Similarity   99.6%;  Pred. No. 0;
  Matches  750;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy      1 MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTREKRGPSGCSGGPNTVYL 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mwalcsllrsaagrtmsqgrtisqaparrerprkdplrhlrtrekrgpsgcsggpntvyl 60

Qy     61 QVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKVARLDNIFLTRMHWSNVGG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 qvvaagsrdsgaalyvfsefnrylfncgegvqrlmqehklkvarldnifltrmhwsnvgg 120

Qy    121 LSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPHSAPEYEDETMTV 180
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    121 lsgmiltlketglpkcvlsgppqlekyleaikifsgplkgielamrphsapeyedetmtv 180

Qy    181 YQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPHGVSQRRGVRDSSLV 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 yqipihseqrrgkhqpwqsperplsrlsperssdsesnenephlphgvsqrrgvrdsslv 240

Qy    241 VAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIAAVKDGKSITHEGREILAEELCTPP 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 vaficklhlkrgnflvlkakemglpvgtaaiapiiaavkdgksithegreilaeelctpp 300

Qy    301 DPGAAFVVVECPDESFIQPICENATFQRYQGKADAPVALVVHMAPASVLVDSRYQQWMER 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 dpgaafvvvecpdesfiqpicenatfqryqgkadapvalvvhmapasvlvdsryqqwmer 360

Qy    361 FGPDTQHLVLNENCASVHNLRSHKIQTQLNLIHPDIFPLLTSFRCKKEGPTLSVPMVQGE 420
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    361 fgpdtqhlvlnencasvhnlrshkiqtqlnlihpdifplltsfrc-kegptlsvpmvqge 419

Qy    421 CLLKYQLRPRREWQRDAIITCNPEEFIVEALQLPNFQQSVQEYRRSAQDGPAPAEKRSQY 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 cllkyqlrprrewqrdaiitcnpeefivealqlpnfqqsvqeyrrsaqdgpapaekrsqy 479

Qy    481 PEIIFLGTGSAIPMKIRNVSATLVNISPDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 peiiflgtgsaipmkirnvsatlvnispdtsllldcgegtfgqlcrhygdqvdrvlgtla 539

Qy    541 AVFVSHLHADHHTGLPSILLQRERALASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLH 600
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    540 avfvshlhadhhtglpsillqreralaslgkplhpllvvapnqlkawlqqyynqcqevlh 599

Qy    601 HISMIPAKCLQEGAEISSPAVERLISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWK 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 hismipakclqegaeisspaverlissllrtcdleefqtclvrhckhafgcalvhtsgwk 659

Qy    661 VVYSGDTMPCEALVRMGKDATLLIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFI 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 vvysgdtmpcealvrmgkdatlliheatledgleeeavekthsttsqaisvgmrmnaefi 719

Qy    721 MLNHFSQRYAKVPLFSPNFSEKVGVAFDHMKVC 753
          |||||||||||||||||||||||||||||||||
Db    720 mlnhfsqryakvplfspnfsekvgvafdhmkvc 752


RESULT   5
AAB07230
ID   AAB07230 standard; protein; 73 AA.
XX
AC   AAB07230;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Murine prostate cancer predisposing protein HPC2.
XX
KW   Mouse; prostate cancer predisposing gene; HPC2; gene therapy;
KW   peptide therapy; drug design.
XX
```

OS Mus musculus.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
DR N-PSDB; AAA60390.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Disclosure; Page 151-152; 157pp; English.
XX
CC The present sequence is the protein produced from the murine prostate
CC cancer predisposing gene HPC2, the human homologue of which is found on
CC chromosome 17p. Some alleles of this gene cause a predisposition to
CC cancer, particularly prostate cancer. This protein and its gene can be
CC used in peptide and gene therapy for cancer patients, as well as being
CC useful as diagnostic tools (both for cancer sufferers and those with a
CC predisposition to the disease) and in the production of cancer drugs.
XX
SQ Sequence 73 AA;

Query Match 1.8%; Score 15; DB 21; Length 73;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      34 KDPLRHLRTREKRGP 48
           |||||||||||||||
Db      28 kdplrhlrtrekrgp 42
```

RESULT 6
AAG18567
ID AAG18567 standard; Protein; 118 AA.
XX
AC AAG18567;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20028.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 10; DB 21; Length 118;
Best Local Similarity 100.0%; Pred. No. 0.07;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 SLLLDCGEGT 520
||||||||||
Db 31 sllldcgegt 40

RESULT 7
AAG18566
ID AAG18566 standard; Protein; 154 AA.
XX
AC AAG18566;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 20027.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.2%; Score 10; DB 21; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 511 SLLLDCGEGT 520
||||||||||
Db 67 sllldcgegt 76

RESULT 8
AAB49964
ID AAB49964 standard; Protein; 838 AA.
XX
AC AAB49964;
XX
DT 08-MAR-2001 (first entry)
XX
DE S. cerevisiae YKR079C protein.
XX
KW Yeast; germination; proliferation; essential gene; antifungal agent;
KW insecticide; herbicide; anti-proliferation drug; cancer; psoriasis;
KW restenosis; YKR081C; YFR003C; YGR277C; YGR278W; YKR071C; YKR079C;
KW YKR083C.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200071161-A1.
XX
PD 30-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13017.
XX
PR 21-MAY-1999; 99US-0315794.
PR 02-SEP-1999; 99US-0389341.
XX
PA (ROSE-) ROSETTA INPHARMATICS INC.
XX
PI Roberts CJ;
XX
DR WPI; 2001-025092/03.
DR N-PSDB; AAC89621.
XX
PT Identifying antifungal compounds which target yeast essential genes
PT comprises use of novel Saccharomyces cerevisiae essential genes
PT YFR003C, YGR277C, YGR278W, YKR0701C, YKR079C or YKR083C or YKR081C -
XX
PS Example 5; Fig 30; 127pp; English.
XX
CC The present invention provides methods of identifying antifungal agents
CC using the coding and protein sequences of several yeast genes. These are
CC essential for the germination and proliferation of Saccharomyces
CC cerevisiae, and include YKR081C, YFR003C, YGR277C, YGR278W, YKR071C,
CC YKR079C and YKR083C. The sequences can also be used to identify compounds
CC for use as herbicides, insecticides and anti-proliferation drugs which
CC can be used in the treatment of cancer, psoriasis and restenosis. This is
CC because they can be used to identify plant, insect and human homologues
CC of the yeast genes.
XX
SQ Sequence 838 AA;

Query Match 1.1%; Score 9; DB 22; Length 838;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 682 LLIHEATLE 690
|||||||||
Db 717 lliheatle 725

RESULT 9
AAB25679
ID AAB25679 standard; Protein; 141 AA.
XX
AC AAB25679;
XX
DT 04-DEC-2000 (first entry)
XX
DE Human secreted protein sequence encoded by gene 15 SEQ ID NO:68.
XX
KW Human; secreted protein; immunosuppressive; immunostimulant; nootropic;
KW antiinflammatory; cardiant; vulnerary; antiulcer; anticonvulsant;
KW antiparkinsonian; neuroprotective; antiviral; antibacterial; cytostatic;
KW antiparasitic; thrombolytic; anticoagulant; antiarteriosclerotic;
KW gene therapy; vaccine; chemotaxis-modulator; angiogenesis-modulator;
KW cancer; immune system disorder; hyperproliferative disorder; infection;
KW cardiovascular disorder; neurological disease; wound healing;
KW chromosome 17.
XX
OS Homo sapiens.
XX
PN WO200043495-A2.
XX
PD 27-JUL-2000.
XX
PF 18-JAN-2000; 2000WO-US00903.
XX
PR 19-JAN-1999; 99US-0116330.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Ebner R, Young PE, Ni J, Moore PA;
PI Komatsoulis G, Birse CE;
XX
DR WPI; 2000-499225/44.
DR N-PSDB; AAA87680.
XX
PT New isolated polynucleotide encoding a secreted protein useful for
PT preventing, treating or ameliorating a medical condition -
XX
PS Claim 11; Page 415; 451pp; English.
XX
CC The polynucleotide sequences given in AAA87666 to AAA87708 encodes the
CC human secreted proteins given in AAB25665 to AAB25755. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC immunostimulant; antiinflammatory; cardiant; vulnerary; antiulcer;
CC nootropic; antiviral; anticonvulsant; antiparkinsonian; neuroprotective;
CC antibacterial; antiparasitic; thrombolytic; anticoagulant;
CC antiarteriosclerotic and cytostatic. The secreted proteins and their
CC polynucleotides can be used in gene therapy and as vaccines,
CC chemotaxis-modulators and angiogenesis- modulators. The human secreted
CC proteins and polynucleotides can be used for diagnosing (the
CC susceptibility to) a pathological condition by determining the presence
CC or absence of a mutation in the polynucleotide or determining the
CC presence or amount of expression of the protein. The polynucleotides and
CC proteins can also be used in the treatment and diagnosis of cancer,

```
CC   diseases of the immune system, hyperproliferative disorders,
CC   cardiovascular disorders and neurological disease. They can also be used
CC   to promote wound healing and to fight infection. AAA87657 to AAA87665 and
CC   AAB25664 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence   141 AA;


  Query Match             1.0%;  Score 8;  DB 21;  Length 141;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     67 SRDSGAAL 74
          ||||||||
Db     46 srdsgaal 53


RESULT  10
AAB43610
ID   AAB43610 standard; Protein; 243 AA.
XX
AC   AAB43610;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated protein sequence SEQ ID NO:1055.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   N-PSDB; AAC77819.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 11; Page 1646-1647; 2352pp; English.
XX
CC   AAC77607 to AAC78448 encode the human cancer associated proteins given
CC   in AAB43398 to AAB44239. The proteins can have activities based on the
CC   tissues and cells the genes are expressed in. Example of activities
CC   include: cytostatic; proliferative; vulnerary; immunomodulator;
CC   antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC   antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC   dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC   nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC   polynucleotides and polypeptides can be used for preventing, treating or
CC   ameliorating medical conditions and diagnosing pathological conditions.
CC   Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC   the present invention may be used to treat immune disorders by activating
CC   or inhibiting the proliferation, differentiation or mobilisation of
CC   immune cells, to treat disorders of haematopoietic cells, autoimmune
CC   disorders, allergic reactions, graft versus host disease and organ
CC   rejection, modulate haemostatic or thrombolytic activity, modulate
CC   inflammation, cancers, cardiovascular disorders, neurological disease and
CC   bacterial or viral infections. The peptides, nucleotides, antibodies,
CC   agonists and antagonists may be also be used in drug screens. AAC78449 to
CC   AAC78457 and AAB44240 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   243 AA;


  Query Match             1.0%;  Score 8;  DB 21;  Length 243;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    298 TPPDPGAA 305
          ||||||||
Db     17 tppdpgaa 24


RESULT  11
AAG75627
ID   AAG75627 standard; Protein; 245 AA.
XX
AC   AAG75627;
XX
DT   03-SEP-2001  (first entry)
XX
DE   Human colon cancer antigen protein SEQ ID NO:6391.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   colorectal carcinoma.
XX
OS   Homo sapiens.
XX
PN   WO200122920-A2.
XX
PD   05-APR-2001.
XX
PF   28-SEP-2000; 2000WO-US26524.
XX
PR   29-SEP-1999;   99US-0157137.
PR   03-NOV-1999;   99US-0163280.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH35032.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS   Claim 11; Page 7872-7873; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where
CC   the proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene
CC   therapy and vaccine production. N and P may be used in the prevention,
CC   diagnosis and treatment of diseases associated with inappropriate P
CC   expression. For example, N and P may be used to treat disorders
CC   associated with decreased expression by rectifying mutations or deletions
CC   in a patient's genome that affect the activity of P by expressing
CC   inactive proteins or to supplement the patients own production of P.
CC   Additionally, N may be used to produce the colon cancer-associated Ps,
CC   by inserting the nucleic acids into a host cell and culturing the cell
CC   to express the proteins. N and P can be used in the prevention, diagnosis
CC   and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC   and AAB77789 represent sequences used in the exemplification of the
```

```
CC   present invention.
CC   N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC   missing at time of publication, meaning no sequences are present for
CC   SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ   Sequence   245 AA;


  Query Match             1.0%;  Score 8;  DB 22;  Length 245;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    298 TPPDPGAA 305
          ||||||||
Db     19 tppdpgaa 26


RESULT  12
AAB42637
ID   AAB42637 standard; Protein; 377 AA.
XX
AC   AAB42637;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2401 polypeptide sequence SEQ ID NO:4802.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; AAC76846.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 3985-3986; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence   377 AA;


  Query Match             1.0%;  Score 8;  DB 21;  Length 377;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     67 SRDSGAAL 74
          ||||||||
Db    283 srdsgaal 290


RESULT  13
AAB85355
ID   AAB85355 standard; Protein; 385 AA.
XX
AC   AAB85355;
XX
DT   17-SEP-2001  (first entry)
XX
DE   Human phosphatase (PP) (clone ID 1269556CD1).
XX
KW   Phosphatase; PP; Alzheimer's disease; Huntington's disease; human;
KW   immune disorder; cancer; nootropic; immunomodulatory; cytostatic;
KW   neuroprotective; anticonvulsant.
XX
OS   Homo sapiens.
XX
PN   WO200153469-A2.
XX
PD   26-JUL-2001.
XX
PF   18-JAN-2001; 2001WO-US02088.
XX
PR   21-JAN-2000; 2000US-0177719.
PR   28-JAN-2000; 2000US-0178988.
PR   25-FEB-2000; 2000US-0184959.
PR   17-MAR-2000; 2000US-0190142.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Bandman O,  Tang YT,  Azimzai Y,  Yue H,  Baughn MR,  Hillman JL;
PI   Lal P,  Wang E,  Gandhi AR,  Policky JL,  Mathur P;
XX
DR   WPI; 2001-451905/48.
DR   N-PSDB; AAH22958.
XX
PT   Novel phosphatases useful for treating diseases associated with
PT   decreased expression of functional phosphatases, e.g., Alzheimer's
PT   disease, Huntington's disease, immune disorders, and cancers -
XX
PS   Claim 1; Page 91-92; 103pp; English.
XX
CC   The invention provides human phosphatases (PP) and polynucleotides
CC   encoding the phosphatases. The polypeptides can be expressed by standard
CC   recombinant methodology. The PP are useful for treating a disease or
```

```
CC   condition associated with decreased expression of functional
CC   phosphatases. Compositions containing agonists or antagonists of PP may
CC   be used to treat a disease associated with decreased expression or
CC   overexpression of PP, respectively. Such diseases may include
CC   Alzheimer's disease, Huntington's disease, immune disorders, and
CC   cancers. The present sequence represents a human phosphatase.
XX
SQ   Sequence   385 AA;


  Query Match             1.0%;  Score 8;  DB 22;  Length 385;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      67 SRDSGAAL 74
           ||||||||
Db     252 srdsgaal 259


RESULT  14
AAG08256
ID   AAG08256 standard; Protein; 394 AA.
XX
AC   AAG08256;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 5719.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
```

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 394;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 513 LLDCGEGT 520
||||||||
Db 57 lldcgegt 64

RESULT 15
AAG08255
ID AAG08255 standard; Protein; 422 AA.
XX
AC AAG08255;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5718.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 422;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     513 LLDCGEGT 520
           ||||||||
Db      85 lldcgegt 92
```

Search completed: March 29, 2002, 10:22:53
Job time: 140 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 09:30:21 ; Search time 1186.93 Seconds
(without alignments)
22461.541 Million cell updates/sec

Title: US-09-988-687-1
Perfect score: 2481
Sequence: 1 atgtgggcgctttgctcgct.....agaaggtcagagcccagtga 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	676	27.2	676	10	BE260495	BE260495 601150702
2	639	25.8	712	10	AU126037	AU126037 AU126037
3	609	24.5	666	10	AU141334	AU141334 AU141334
4	601	24.2	762	10	AU124671	AU124671 AU124671
5	600	24.2	791	11	BG822529	BG822529 602725613
6	597	24.1	944	11	BF969043	BF969043 602269966
7	590	23.8	664	10	BE383336	BE383336 601298249
8	590	23.8	692	10	BE382353	BE382353 601298656
9	561	22.6	920	11	BG386090	BG386090 602455264
10	556	22.4	890	11	BG756043	BG756043 602716533
11	553	22.3	553	11	BF434169	BF434169 7o99f04.x
12	543	21.9	827	11	BG396395	BG396395 602459323
13	530	21.4	614	11	BG335963	BG335963 602404708
14	528	21.3	761	11	BE795820	BE795820 601590856
15	526	21.2	526	11	BF196304	BF196304 7n68a10.x
16	525	21.2	696	10	AU138595	AU138595 AU138595
17	521	21.0	688	11	BE794311	BE794311 601591442
18	519	20.9	960	11	BF797306	BF797306 602256926
19	507	20.4	670	10	AU143668	AU143668 AU143668
20	506	20.4	506	11	BF057321	BF057321 7k19c01.x
21	505	20.4	771	11	BG766403	BG766403 602739146
22	497	20.0	953	11	BF794577	BF794577 602255843
23	491	19.8	884	11	BG519751	BG519751 602578902
24	478	19.3	1012	11	BG386348	BG386348 602455550
25	477	19.2	657	10	AU138795	AU138795 AU138795
26	477	19.2	735	11	BG719016	BG719016 602699144
27	475	19.1	915	11	BF311926	BF311926 601897767
28	469	18.9	883	11	BG746363	BG746363 602703625
29	465	18.7	823	10	BE619259	BE619259 601473130
30	458	18.5	531	10	AA243700	AA243700 zr68g08.s
31	453	18.3	611	11	BG471513	BG471513 602513052
32	453	18.3	760	11	BG751942	BG751942 602732110
33	451	18.2	451	10	AI200296	AI200296 qf86b12.x
34	450	18.1	812	11	BE867512	BE867512 601443010
35	447	18.0	498	11	BF433563	BF433563 7q55b07.x
36	447	18.0	938	10	BE260626	BE260626 601146116
37	442	17.8	493	10	AW575677	AW575677 UI-HF-BM0
38	438	17.7	612	10	BE304720	BE304720 601106236
39	436	17.6	493	10	AA455121	AA455121 zx78c04.s
40	421	17.0	421	10	AI468143	AI468143 tf92g05.x
41	421	17.0	536	10	AW206103	AW206103 UI-H-BI1-
42	406	16.4	452	11	BG327066	BG327066 602426274
43	405	16.3	457	11	BF309436	BF309436 601892128
44	404	16.3	404	10	AW304130	AW304130 xs13e05.x
45	402	16.2	474	10	AA291670	AA291670 zt37d04.s

ALIGNMENTS

RESULT 1
BE260495
LOCUS BE260495 676 bp mRNA EST 26-OCT-2000
DEFINITION 601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5', mRNA sequence.
ACCESSION BE260495
VERSION BE260495.1 GI:9131807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 676)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM176 row: d column: 01
High quality sequence stop: 672.
FEATURES Location/Qualifiers
source 1. .676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3503184"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 154 a 207 c 176 g 139 t
ORIGIN

Query Match 27.2%; Score 676; DB 10; Length 676;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1093 acccagcacttggtcctgaatgagaactgtgcctcagttcacaaccttcgcagccacaag 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAG 60

Qy   1153 attcaaacccagctcaacctcatccacccggacatcttccccctgctcaccagtttccgc 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCGC 120

Qy   1213 tgtaagaaggagggccccaccctcagtgtgcccatggttcagggtgaatgcctcctcaag 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAG 180

Qy   1273 taccagctccgtcccaggagggagtggcagagggatgccattattacttgcaatcctgag 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAG 240

Qy   1333 gaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtgcaggagtacagg 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 300

Qy   1393 aggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtacccagaaatcatc 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATC 360

Qy   1453 ttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagtgccacacttgtc 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTC 420

Qy   1513 aacataagccccgacacgtctctgctactggactgtggtgagggcacatttgggcagctg 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTG 480

Qy   1573 tgccgtcattacggagaccaggtggacagggtcctgggcaccctggctgctgtgtttgtg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTG 540

Qy   1633 tcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctgcagagagaacgc 1692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGC 600

Qy   1693 gccttggcatctttgggaaagccgcttcacccttgctggtggttgcccccaaccagctc 1752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGCTC 660

Qy   1753 aaagcctggctccagc 1768
          ||||||||||||||||
Db    661 AAAGCCTGGCTCCAGC 676
```

RESULT 2
AU126037
LOCUS AU126037 712 bp mRNA EST 23-OCT-2000
DEFINITION AU126037 NT2RM4 Homo sapiens cDNA clone NT2RM4002610 5', mRNA sequence.
ACCESSION AU126037
VERSION AU126037.1 GI:10950753
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 712)
AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and Isogai,T.
TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES Location/Qualifiers
source 1. .712
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4002610"
/clone_lib="NT2RM4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"
BASE COUNT 173 a 198 c 204 g 134 t 3 others
ORIGIN

Query Match 25.8%; Score 639; DB 10; Length 712;
Best Local Similarity 100.0%; Pred. No. 7.1e-297;
Matches 639; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     22 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 81

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 141

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 201

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 261

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 321

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 381

Qy    361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 441

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 501

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    502 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 561

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    562 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 621

Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttca 639
          |||||||||||||||||||||||||||||||||||||||
Db    622 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCA 660


RESULT  3
AU141334
LOCUS       AU141334      666 bp    mRNA            EST       25-OCT-2000
DEFINITION  AU141334 THYRO1 Homo sapiens cDNA clone THYRO1000421 5', mRNA
            sequence.
ACCESSION   AU141334
VERSION     AU141334.1  GI:11002855
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 666)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .666
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="THYRO1000421"
                     /clone_lib="THYRO1"
                     /tissue_type="thyroid gland"
                     /note="Vector: pME18SFL3"
BASE COUNT      158 a    186 c    190 g    129 t      3 others
ORIGIN


  Query Match             24.5%;  Score 609;  DB 10;  Length 666;
  Best Local Similarity   100.0%;  Pred. No. 2e-282;
  Matches  609;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     29 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 88

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     89 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 148

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    149 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 208

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    209 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 268

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    269 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 328

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    329 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 388

Qy    361 ttaagtggaatgattcttacttttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    389 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 448

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    449 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 508

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    509 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 568

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    569 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 628

Qy    601 gaaaggcct 609
          |||||||||
Db    629 GAAAGGCCT 637


RESULT  4
AU124671
LOCUS       AU124671      762 bp    mRNA            EST       23-OCT-2000
DEFINITION  AU124671 NT2RM4 Homo sapiens cDNA clone NT2RM4000375 5', mRNA
            sequence.
ACCESSION   AU124671
VERSION     AU124671.1  GI:10949387
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
            ,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3951
            Fax: 81-438-52-3952
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .762
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NT2RM4000375"
                     /clone_lib="NT2RM4"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
BASE COUNT      180 a    212 c    202 g    165 t      3 others
ORIGIN
```

```
  Query Match             24.2%;  Score 601;  DB 10;  Length 762;
  Best Local Similarity   99.8%;  Pred. No. 1.5e-278;
  Matches  651;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    852 cactcatgaaggaagagagattttggctgaagagctgtgtactcctccagatcctggtgc 911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCAGATCCTGGTGC 60

Qy    912 tgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatctgtgagaatgc 971
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTGTGAGAATGC 120

Qy    972 caccttttcagaggtaccaaggaaaggcagatgcccccgtggccttggtggttcacatggc 1031
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTGGTTCACATGGC 180

Qy   1032 cccagcatctgtgcttgtggacagcaggtaccagcagtggatggagaggtttgggcctga 1091
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGA 240

Qy   1092 cacccagcacttggtcctgaatgagaactgtgcctcagttcacaaccttcgcagccacaa 1151
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAA 300

Qy   1152 gattcaaacccagctcaacctcatccacccggacatcttcccccctgctcaccagtttccg 1211
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCG 360

Qy   1212 ctgtaagaaggagggccccacccctcagtgtgcccatggttcagggtgaatgcctcctcaa 1271
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAA 420

Qy   1272 gtaccagctccgtcccaggagggagtggcagagggatgccattattacttgcaatcctga 1331
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGA 480

Qy   1332 ggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtgcaggagtacag 1391
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    481 GGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCATCAGAGCGTGCAGGAGTACAG 540

Qy   1392 gaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtacccagaaatcat 1451
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCAT 600

Qy   1452 cttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagtgcc 1503
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCC 652


RESULT   5
BG822529
LOCUS       BG822529      791 bp    mRNA            EST       22-MAY-2001
DEFINITION  602725613F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4865354 5',
            mRNA sequence.
ACCESSION   BG822529
VERSION     BG822529.1  GI:14170116
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 791)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1730  row: i  column: 03
            High quality sequence stop: 789.
FEATURES             Location/Qualifiers
     source          1. .791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4865354"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      182 a    224 c    216 g    169 t
ORIGIN


  Query Match             24.2%;  Score 600;  DB 11;  Length 791;
  Best Local Similarity   99.7%;  Pred. No. 4.5e-278;
  Matches  770;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy    916 tttgtggtggtagaatgtccagatgaaagcttcattcaacccatctgtgagaatgccacc 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 TTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTGTGAGAATGCCACC 80

Qy    976 tttcagaggtaccaaggaaaggcagatgcccccgtggccttggtggttcacatggcccca 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 TTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCA 140

Qy   1036 gcatctgtgcttgtggacagcaggtaccagcagtggatggagaggtttgggcctgacacc 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 GCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACC 200

Qy   1096 cagcacttggtcctgaatgagaactgtgcctcagttcacaaccttcgcagccacaagatt 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 CAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAGATT 260

Qy   1156 caaacccagctcaacctcatccacccggacatcttcccccctgctcaccagtttccgctgt 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 CAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCGCTGT 320

Qy   1216 aagaaggagggccccacccctcagtgtgcccatggttcagggtgaatgcctcctcaagtac 1275
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 AAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAGTAC 380

Qy   1276 cagctccgtcccaggagggagtggcagagggatgccattattacttgcaatcctgaggaa 1335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 CAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAGGAA 440

Qy   1336 ttcatagttgaggcgctgcagcttcccaacttccagcagagcgtgcaggagtacaggagg 1395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 TTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGG 500

Qy   1396 agtgcgcaggacggcccagccccagcagagaaaagaagtcagtacccagaaatcatcttc 1455
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    501 AGTGCGCAGGACGG-CCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATCTTC 559

Qy   1456 cttggaacagggtctgccatcccgatgaagattcgaaatgtcagtgccacacttgtcaac 1515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    560 CTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTCAAC 619

Qy   1516 ataagccccgacacgtctctgctactggactgtggtgagggcacatttgggcagctgtgc 1575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 ATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTGTGC 679

Qy   1576 cgtcattacggagaccaggtggacagggtcctgggcaccctggctgctgtgtttgtgtcc 1635
```

```
Db    680 CGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGTTGCTGTGTTTGTGTCC 739

Qy   1636 cacctgcacgcagatcaccacacgggcttgccaagtatcttgctgcagagag 1687
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    740 CACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTGCAGAGAG 791


RESULT   6
BF969043
LOCUS       BF969043      944 bp    mRNA            EST       22-JAN-2001
DEFINITION  602269966F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357985 5',
            mRNA sequence.
ACCESSION   BF969043
VERSION     BF969043.1  GI:12336258
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 944)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9995  row: d  column: 18
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .944
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4357985"
                     /clone_lib="NIH_MGC_84"
                     /tissue_type="adrenal cortex carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
                     NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
                     primed. Average insert size 1.229 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      200 a    266 c    264 g    214 t
ORIGIN


  Query Match             24.1%;  Score 597;  DB 11;  Length 944;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-276;
  Matches  597;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    796 gttgggacagctgccatcgctcccatcattgctgctgtcaaggacgggaaaagcatcact 855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGACGGGAAAAGCATCACT 60

Qy    856 catgaaggaagagagattttggctgaagagctgtgtactcctccagatcctggtgctgct 915
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCAGATCCTGGTGCTGCT 120

Qy    916 tttgtggtggtagaatgtccagatgaaagcttcattcaacccatctgtgagaatgccacc 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTGTGAGAATGCCACC 180

Qy    976 tttcagaggtaccaaggaaaggcagatgcccccgtggccttggtggttcacatggcccca 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCA 240

Qy   1036 gcatctgtgcttgtggacagcaggtaccagcagtggatggagaggtttgggcctgacacc 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACC 300

Qy   1096 cagcacttggtcctgaatgagaactgtgcctcagttcacaaccttcgcagccacaagatt 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAGATT 360

Qy   1156 caaacccagctcaacctcatccacccggacatcttccccctgctcaccagtttccgctgt 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCGCTGT 420

Qy   1216 aagaaggagggccccacctcagtgtgcccatggttcagggtgaatgcctcctcaagtac 1275
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAGTAC 480

Qy   1276 cagctccgtcccaggagggagtggcagagggatgccattattacttgcaatcctgaggaa 1335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAGGAA 540

Qy   1336 ttcatagttgaggcgctgcagcttcccaacttccagcagagcgtgcaggagtacagg 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 597


RESULT   7
BE383336
LOCUS       BE383336      664 bp    mRNA            EST       21-JUL-2000
DEFINITION  601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
            mRNA sequence.
ACCESSION   BE383336
VERSION     BE383336.1  GI:9328701
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM313  row: a  column: 13
            High quality sequence stop: 662.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3628308"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      154 a    199 c    174 g    137 t
ORIGIN


  Query Match             23.8%;  Score 590;  DB 10;  Length 664;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-273;
  Matches  590;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy   1093 acccagcacttggtcctgaatgagaactgtgcctcagttcacaaccttcgcagccacaag 1152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAG 60

Qy   1153 attcaaacccagctcaacctcatccacccggacatcttccccctgctcaccagtttccgc 1212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCGC 120

Qy   1213 tgtaagaaggagggccccaccctcagtgtgcccatggttcagggtgaatgcctcctcaag 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAG 180

Qy   1273 taccagctccgtcccaggagggagtggcagagggatgccattattacttgcaatcctgag 1332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAG 240

Qy   1333 gaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtgcaggagtacagg 1392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 300

Qy   1393 aggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtacccagaaatcatc 1452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATC 360

Qy   1453 ttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagtgccacacttgtc 1512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTC 420

Qy   1513 aacataagccccgacacgtctctgctactggactgtggtgagggcacatttgggcagctg 1572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTG 480

Qy   1573 tgccgtcattacggagaccaggtggacagggtcctgggcaccctggctgctgtgtttgtg 1632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTG 540

Qy   1633 tcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctgca 1682
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTGCA 590


RESULT   8
BE382353
LOCUS       BE382353      692 bp    mRNA            EST       21-JUL-2000
DEFINITION  601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
            mRNA sequence.
ACCESSION   BE382353
VERSION     BE382353.1  GI:9327718
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM314  row: o  column: 13
            High quality sequence stop: 600.
FEATURES             Location/Qualifiers
     source          1. .692
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3629028"
                     /clone_lib="NIH_MGC_19"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      152 a    193 c    197 g    150 t
ORIGIN


  Query Match             23.8%;  Score 590;  DB 10;  Length 692;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-273;
  Matches  590;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1475 tcccgatgaagattcgaaatgtcagtgccacacttgtcaacataagccccgacacgtctc 1534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTC 60

Qy   1535 tgctactggactgtggtgagggcacatttgggcagctgtgccgtcattacggagaccagg 1594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTGTGCCGTCATTACGGAGACCAGG 120

Qy   1595 tggacagggtcctgggcaccctggctgctgtgtttgtgtcccacctgcacgcagatcacc 1654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGACAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGATCACC 180

Qy   1655 acacgggcttgccaagtatcttgctgcagagagaacgcgccttggcatctttgggaaagc 1714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACACGGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGC 240

Qy   1715 cgcttcacccttgctggtggttgcccccaaccagctcaaagcctggctccagcagtacc 1774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGCTTCACCCTTTGCTGGTGGTTGCCCCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACC 300

Qy   1775 acaaccagtgccaggaggtcctgcaccacatcagtatgattcctgccaaatgccttcagg 1834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACAACCAGTGCCAGGAGGTCCTGCACCACATCAGTATGATTCCTGCCAAATGCCTTCAGG 360

Qy   1835 aaggggctgagatctccagtcctgcagtggaaagattgatcagttcgctgttgcgaacat 1894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGGGCTGAGATCTCCAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACAT 420

Qy   1895 gtgatttggaagagtttcagacctgtctggtgcggcactgcaagcatgcgtttggctgtg 1954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GTGATTTGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTG 480

Qy   1955 cgctggtgcacacctctggctggaaagtggtctattccggggacaccatgccctgcgagg 2014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CGCTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACACCATGCCCTGCGAGG 540

Qy   2015 ctctggtccggatggggaaagatgccaccctcctgatacatgaagccacc 2064
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCTGGTCCGGATGGGGAAAGATGCCACCCTCCTGATACATGAAGCCACC 590


RESULT   9
BG386090
LOCUS       BG386090      920 bp    mRNA            EST       12-MAR-2001
DEFINITION  602455264F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583679 5',
            mRNA sequence.
ACCESSION   BG386090
VERSION     BG386090.1  GI:13279536
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 920)
```

```
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: NIH Intramural Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1307  row: p  column: 16
            High quality sequence stop: 736.
FEATURES             Location/Qualifiers
     source          1. .920
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4583679"
                     /clone_lib="NIH_MGC_15"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)"
BASE COUNT      196 a    261 c    272 g    190 t      1 others
ORIGIN


  Query Match             22.6%;  Score 561;  DB 11;  Length 920;
  Best Local Similarity   99.9%;  Pred. No. 2.8e-259;
  Matches  681;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy    916 tttgtggtggtagaatgtccagatgaaagcttcattcaacccatctgtgagaatgccacc 975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 TTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATCTGTGAGAATGCCACC 80

Qy    976 tttcagaggtaccaaggaaaggcagatgcccccgtggccttggtggttcacatggccccca 1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 TTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTGGTTCACATGGCCCCA 140

Qy   1036 gcatctgtgcttgtggacagcaggtaccagcagtggatggagaggtttgggcctgacacc 1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 GCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCTGACACC 200

Qy   1096 cagcacttggtcctgaatgagaactgtgcctcagttcacaaccttcgcagccacaagatt 1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 CAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAGATT 260

Qy   1156 caaacccagctcaacctcatccacccggacatcttccccctgctcaccagtttccgctgt 1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 CAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTTCCGCTGT 320

Qy   1216 aagaaggagggccccacctcagtgtgcccatggttcagggtgaatgcctcctcaagtac 1275
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 AAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAATGCCTCCTCAAGTAC 380

Qy   1276 cagctccgtcccaggagggagtggcagagggatgccattattacttgcaatcctgaggaa 1335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 CAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACTTGCAATCCTGAGGAA 440

Qy   1336 ttcatagttgaggcgctgcagcttcccaacttccagcagagcgtgcaggagtacaggagg 1395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 TTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGGAGG 500

Qy   1396 agtgcgcaggacggcccagccccagcagagaaaagaagtcagtacccagaaatcatcttc 1455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 AGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTACCCAGAAATCATCTTC 560

Qy   1456 cttggaacagggtctgccatcccgatgaagattcgaaatgtcagtgccacacttgtcaac 1515
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    561 CTTGGAACAGGGTCTG-CATCCCGATGAAGATTCGAAATGTCAGTGCCACACTTGTCAAC 619

Qy   1516 ataagccccgacacgtctctgctactggactgtggtgagggcacatttgggcagctgtgc 1575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    620 ATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTGTGC 679

Qy   1576 cgtcattacggagaccaggtgg 1597
          |||||||||||||||||||||
Db    680 CGTCATTACGGAGACCAGGTGG 701


RESULT  10
BG756043
LOCUS       BG756043      890 bp    mRNA            EST       15-MAY-2001
DEFINITION  602716533F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4856360 5',
            mRNA sequence.
ACCESSION   BG756043
VERSION     BG756043.1  GI:14066696
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1707  row: b  column: 09
            High quality sequence stop: 878.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4856360"
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      209 a    247 c    275 g    159 t
ORIGIN


  Query Match             22.4%;  Score 556;  DB 11;  Length 890;
  Best Local Similarity   100.0%;  Pred. No. 7.2e-257;
  Matches  556;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1926 gcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaagtggt 1985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGGT 60

Qy   1986 ctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgccaccct 2045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCCACCCT 120
```

```
Qy   2046 cctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaagacaca 2105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCTGATACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAGACACA 180

Qy   2106 cagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcattatgct 2165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATTATGCT 240

Qy   2166 gaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagcgagaa 2225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGCGAGAA 300

Qy   2226 agtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatgcccaa 2285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGAGACTTTCCAACAATGCCCAA 360

Qy   2286 gctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggagcgcag 2345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTGATTCCCCCACTGAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAGCGCAG 420

Qy   2346 ggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggcaggcgg 2405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGAGAAGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCAGGCGG 480

Qy   2406 cctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggccaagaa 2465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGAGCCACAGGCCAAGAA 540

Qy   2466 ggtcagagcccagtga 2481
          ||||||||||||||||
Db    541 GGTCAGAGCCCAGTGA 556
```

```
RESULT  11
BF434169
LOCUS       BF434169      553 bp    mRNA            EST       29-NOV-2000
DEFINITION  7o99f04.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3644670 3'
            similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION   BF434169
VERSION     BF434169.1  GI:11446441
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40UP from Gibco
            High quality sequence stop: 507.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3644670"
                     /clone_lib="NCI_CGAP_Ov18"
                     /tissue_type="fibrotheoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCGACATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not
                     I and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      143 a    150 c    153 g    107 t
ORIGIN
```

```
  Query Match             22.3%;  Score 553;  DB 11;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2e-255;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     91 cggccgcgcaaggacccgctgcggcacctgcgcacgcgagagaagcgcggaccgtcgggg 150
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGCCGCGCAAGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGGG 60

Qy    151 tgctccggcggcccaaacaccgtgtacctgcaggtggtggcagcgggtagccgggactcg 210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAGCCGGGACTCG 120

Qy    211 ggcgccgcgctctacgtcttctccgagttcaaccggtatctcttcaactgtggagaaggc 270
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATCTCTTCAACTGTGGAGAAGGC 180

Qy    271 gttcagagactcatgcaggagcacaagttaaaggttgctcgcctggacaacatattcctg 330
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTTCAGAGACTCATGCAGGAGCACAAGTTAAAGGTTGCTCGCCTGGACAACATATTCCTG 240

Qy    331 acacgaatgcactggtctaatgttgggggcttaagtggaatgattcttactttaaaggaa 390
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ACACGAATGCACTGGTCTAATGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGAA 300

Qy    391 accgggcttccaaagtgtgtactttctggacctccacaactggaaaaatacctcgaagca 450
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ACCGGGCTTCCAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGCA 360

Qy    451 atcaaaatattttctggtccattgaaaggaatagaactggctgtgcggccccactctgcc 510
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCCCCACTCTGCC 420

Qy    511 ccagaatacgaggatgaaaccatgacagtttaccagatccccatacacagtgaacagagg 570
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCCCCATACACAGTGAACAGAGG 480

Qy    571 aggggaaagcaccaaccatggcagagtccagaaaggcctctcagcaggctcagtccagag 630
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGGGGAAAGCACCAACCATGGCAGAGTCCAGAAAGGCCTCTCAGCAGGCTCAGTCCAGAG 540

Qy    631 cgatcttcagact 643
          |||||||||||||
Db    541 CGATCTTCAGACT 553
```

```
RESULT  12
BG396395
LOCUS       BG396395      827 bp    mRNA            EST       12-MAR-2001
DEFINITION  602459323F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4581700 5',
            mRNA sequence.
ACCESSION   BG396395
VERSION     BG396395.1  GI:13289941
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
```

```
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1302  row: n  column: 05
            High quality sequence stop: 779.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4581700"
                     /clone_lib="NIH_MGC_16"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      205 a    217 c    226 g    179 t
ORIGIN


  Query Match             21.9%;  Score 543;  DB 11;  Length 827;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-250;
  Matches  543;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     707 actcttccctggtcgtagctttcatctgtaagcttcacttaaagagaggaaacttcttgg 766
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      22 ACTCTTCCCTGGTCGTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGG 81

Qy     767 tgctcaaagcaaaggagatgggcctcccagttgggacagctgccatcgctcccatcattg 826
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82 TGCTCAAAGCAAAGGAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCATTG 141

Qy     827 ctgctgtcaaggacgggaaaagcatcactcatgaaggaagagagattttggctgaagagc 886
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     142 CTGCTGTCAAGGACGGGAAAAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGC 201

Qy     887 tgtgtactcctccagatcctggtgctgcttttgtggtggtagaatgtccagatgaaagct 946
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     202 TGTGTACTCCTCCAGATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCT 261

Qy     947 tcattcaacccatctgtgagaatgccacctttcagaggtaccaaggaaaggcagatgccc 1006
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     262 TCATTCAACCCATCTGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCC 321

Qy    1007 ccgtggccttggtggttcacatggccccagcatctgtgcttgtggacagcaggtaccagc 1066
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     322 CCGTGGCCTTGGTGGTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGC 381

Qy    1067 agtggatggagaggtttgggcctgacacccagcacttggtcctgaatgagaactgtgcct 1126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     382 AGTGGATGGAGAGGTTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCT 441

Qy    1127 cagttcacaaccttcgcagccacaagattcaaacccagctcaacctcatccacccggaca 1186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     442 CAGTTCACAACCTTCGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACA 501

Qy    1187 tcttccccctgctcaccagtttccgctgtaagaaggagggccccaccctcagtgtgccca 1246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     502 TCTTCCCCCTGCTCACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCA 561

Qy    1247 tgg 1249
           |||
Db     562 TGG 564
```

```
RESULT  13
BG335963
LOCUS       BG335963      614 bp    mRNA            EST       27-FEB-2001
DEFINITION  602404708F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542553 5',
            mRNA sequence.
ACCESSION   BG335963
VERSION     BG335963.1  GI:13142401
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1222  row: o  column: 02
            High quality sequence stop: 614.
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4542553"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      135 a    169 c    178 g    132 t
ORIGIN


  Query Match             21.4%;  Score 530;  DB 11;  Length 614;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-244;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1505 cacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcacatttg 1564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 CACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTG 61

Qy    1565 ggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggctgctg 1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 GGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCTGCTG 121

Qy    1625 tgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctgcaga 1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 TGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTGCAGA 181

Qy    1685 gagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgccccca 1744
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTGCTGGTGGTTGCCCCCA 241

Qy    1745 accagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcaccaca 1804
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 ACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCACCACA 301

Qy    1805 tcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgcagtgg 1864
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 TCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCAGTGG 361

Qy   1865 aaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgtctgg 1924
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 AAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGTCTGG 421

Qy   1925 tgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaagtgg 1984
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 TGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAAGTGG 481

Qy   1985 tctattccggggacaccatgccctgcgaggctctggtccggatggggaaa 2034
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 TCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAA 531


RESULT  14
BE795820
LOCUS       BE795820      761 bp    mRNA            EST       20-SEP-2000
DEFINITION  601590856F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
            mRNA sequence.
ACCESSION   BE795820
VERSION     BE795820.1  GI:10217018
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 761)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM802  row: p  column: 14
            High quality sequence stop: 761.
FEATURES             Location/Qualifiers
     source          1. .761
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3945085"
                     /clone_lib="NIH_MGC_7"
                     /tissue_type="small cell carcinoma"
                     /cell_line="MGC3"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      172 a    206 c    219 g    164 t
ORIGIN


  Query Match             21.3%;  Score 528;  DB 11;  Length 761;
  Best Local Similarity   99.7%;  Pred. No. 2.3e-243;
  Matches  698;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy   1479 gatgaagattcgaaatgtcagtgccacacttgtcaacataagccccgacacgtctctgct 1538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GATGAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCT 60

Qy   1539 actggactgtggtgagggcacatttgggcagctgtgccgtcattacggagaccaggtgga 1598
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||

Db     61 ACTGGACTGTGGTGAGGGCACGTTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGA 120

Qy   1599 cagggtcctgggcaccctggctgctgtgtttgtgtcccacctgcacgcagatcaccacac 1658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACAC 180

Qy   1659 gggcttgccaagtatcttgctgcagagagaacgcgccttggcatctttgggaaagccgct 1718
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGGCTTGCCAAGTATCTTGCTGCAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCT 240

Qy   1719 tcaccctttgctggtggttgccccccaaccagctcaaagcctggctccagcagtaccacaa 1778
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCACCCTTTGCTGGTGGTTGCCCCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAA 300

Qy   1779 ccagtgccaggaggtcctgcaccacatcagtatgattcctgccaaatgccttcaggaagg 1838
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCAGTGCCAGGAGGTCCTGCACCACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGG 360

Qy   1839 ggctgagatctccagtcctgcagtggaaagattgatcagttcgctgttgcgaacatgtga 1898
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGCTGAGATCTCCAGTCCTGCAGTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGA 420

Qy   1899 tttggaagagtttcagacctgtctggtgcggcactgcaagcatgcgtttggctgtgcgct 1958
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 -TTGGAAGAGTTTCAGACCTGTCTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCT 479

Qy   1959 ggtgcacacctctggctggaaagtggtctattccggggacaccatgccctgcgaggctct 2018
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 GGTGCACACCTCTGGCTGGAAAGTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCT 539

Qy   2019 ggtccggatggggaaagatgccaccctcctgatacatgaagccaccctggaagatggttt 2078
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 GGTCCGGATGGGGAAAGATGCCACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTT 599

Qy   2079 ggaagaggaagcagtggaaaagacacacagcacaacgtcccaagccatcagcgtggggat 2138
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 GGAAGAGGAAGCAGTGGAAAAGACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGAT 659

Qy   2139 gcggatgaacgcggagttcattatgctgaaccacttcagc 2178
          ||||||||||||||||||||||||||||||||||||||||
Db    660 GCGGATGAACGCGGAGTTCATTATGCTGAACCACTTCAGC 699


RESULT  15                                                              -
BF196304
LOCUS       BF196304      526 bp    mRNA            EST       03-NOV-2000
DEFINITION  7n68a10.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3569586 3'
            similar to TR:Q9V5J4 Q9V5J4 CG3298 PROTEIN. ;, mRNA sequence.
ACCESSION   BF196304
VERSION     BF196304.1  GI:11084107
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 526)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento
            Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The
            I.M.A.G.E. Consortium DNA Sequencing by: Washington University
            Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            High quality sequence stop: 491.
FEATURES             Location/Qualifiers
     source          1. .526
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3569586"
/clone_lib="NCI_CGAP_Ov18"
/tissue_type="fibrotheoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGCGACATTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 136 a 142 c 147 g 101 t
ORIGIN

Query Match 21.2%; Score 526; DB 11; Length 526;
Best Local Similarity 100.0%; Pred. No. 2.1e-242;
Matches 526; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      90 gcggccgcgcaaggacccgctgcggcacctgcgcacgcgagagaagcgcggaccgtcggg 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCGGCCGCGCAAGGACCCGCTGCGGCACCTGCGCACGCGAGAGAAGCGCGGACCGTCGGG 60

Qy     150 gtgctccggcggcccaaacaccgtgtacctgcaggtggtggcagcgggtagccgggactc 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTGCTCCGGCGGCCCAAACACCGTGTACCTGCAGGTGGTGGCAGCGGGTAGCCGGGACTC 120

Qy     210 gggcgccgcgctctacgtcttctccgagttcaaccggtatctcttcaactgtggagaagg 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGGCGCCGCGCTCTACGTCTTCTCCGAGTTCAACCGGTATCTCTTCAACTGTGGAGAAGG 180

Qy     270 cgttcagagactcatgcaggagcacaagttaaaggttgctcgcctggacaacatattcct 329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGTTCAGAGACTCATGCAGGAGCACAAGTTAAAGGTTGCTCGCCTGGACAACATATTCCT 240

Qy     330 gacacgaatgcactggtctaatgttgggggcttaagtggaatgattcttactttaaagga 389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GACACGAATGCACTGGTCTAATGTTGGGGGCTTAAGTGGAATGATTCTTACTTTAAAGGA 300

Qy     390 aaccgggcttccaaagtgtgtactttctggacctccacaactggaaaaatacctcgaagc 449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AACCGGGCTTCCAAAGTGTGTACTTTCTGGACCTCCACAACTGGAAAAATACCTCGAAGC 360

Qy     450 aatcaaaatattttctggtccattgaaaggaatagaactggctgtgcggccccactctgc 509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGCTGTGCGGCCCCACTCTGC 420

Qy     510 cccagaatacgaggatgaaaccatgacagtttaccagatccccatacacagtgaacagag 569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCCAGAATACGAGGATGAAACCATGACAGTTTACCAGATCCCCATACACAGTGAACAGAG 480

Qy     570 gaggggaaagcaccaaccatggcagagtccagaaaggcctctcagc 615
           |||||||||||||||||||||||||||||||||||||||||||||
Db     481 GAGGGGAAAGCACCAACCATGGCAGAGTCCAGAAAGGCCTCTCAGC 526
```

Search completed: April 1, 2002, 10:37:18
Job time: 4017 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 09:33:06 ; Search time 132.23 Seconds
(without alignments)
16085.789 Million cell updates/sec

Title: US-09-988-687-1
Perfect score: 2481
Sequence: 1 atgtgggcgctttgctcgct.....agaaggtcagagcccagtga 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_1101:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2481	100.0	2958	21	AAA58453	Human prostate can
	2	2274	91.7	2478	21	AAA52810	Human sulphatase G
	3	2022	81.5	2976	22	AAH14250	Human cDNA sequenc
	4	1600	64.5	2546	21	AAC76445	Human ORFX ORF2000
	5	364	14.7	584	22	AAH05835	Human cDNA clone (
	6	247	10.0	26664	21	AAA60207	Human prostate can
	7	237	9.6	238	21	AAZ80231	Human colon cancer
	8	26	1.0	326	21	AAA60390	Murine prostate ca
	9	23	0.9	23	21	AAA60373	Human prostate can
c	10	21	0.8	143	21	AAC24564	Human secreted pro
c	11	20	0.8	38	21	AAA60275	Human HPC2 cDNA ex
	12	20	0.8	308	21	AAF09297	Fusarium venenatum
	13	20	0.8	457	21	AAC57035	Pinus radiata tran
c	14	19	0.8	19	21	AAA60211	Human HPC2 cDNA 5'
	15	19	0.8	19	21	AAA60229	Human HPC2 cDNA se
c	16	19	0.8	19	21	AAA60230	Human HPC2 cDNA se
	17	19	0.8	19	21	AAA60233	Human HPC2 cDNA se
	18	19	0.8	19	21	AAA60235	Human HPC2 cDNA se
	19	19	0.8	19	21	AAA60364	Human HPC2 cDNA ex
c	20	19	0.8	29	21	AAA60214	Human HPC2 cDNA ex
	21	19	0.8	37	21	AAA60362	Human HPC2 cDNA ex
c	22	19	0.8	38	21	AAA60307	Human HPC2 cDNA ex
c	23	19	0.8	526	20	AAV88751	EST clone HK189.
	24	19	0.8	7642	20	AAV82020	Moraxella catarrha
	25	18	0.7	18	21	AAA60219	Human HPC2 cDNA se
c	26	18	0.7	18	21	AAA60220	Human HPC2 cDNA se
	27	18	0.7	18	21	AAA60221	Human HPC2 cDNA se
c	28	18	0.7	18	21	AAA60222	Human HPC2 cDNA se
	29	18	0.7	18	21	AAA60223	Human HPC2 cDNA se
c	30	18	0.7	18	21	AAA60224	Human HPC2 cDNA se
	31	18	0.7	18	21	AAA60225	Human HPC2 cDNA se
c	32	18	0.7	18	21	AAA60226	Human HPC2 cDNA se
	33	18	0.7	18	21	AAA60227	Human HPC2 cDNA se
c	34	18	0.7	18	21	AAA60228	Human HPC2 cDNA se
c	35	18	0.7	18	21	AAA60232	Human HPC2 cDNA se
	36	18	0.7	18	21	AAA60236	Human HPC2 cDNA se
	37	18	0.7	18	21	AAA60237	Human HPC2 cDNA se
	38	18	0.7	18	21	AAA60238	Human HPC2 cDNA se
	39	18	0.7	18	21	AAA60240	Human HPC2 cDNA se
c	40	18	0.7	18	21	AAA60243	Human HPC2 cDNA se
	41	18	0.7	20	22	AAS11239	Human ANK gene PCR
c	42	18	0.7	21	22	AAS11259	Human ANK gene PCR
	43	18	0.7	45	20	AAZ33905	Human PRO274 hybri
	44	18	0.7	45	21	AAC78611	Human PRO274 hybri
c	45	18	0.7	177	20	AAX18066	Coding sequence fo

ALIGNMENTS

RESULT 1
AAA58453
ID AAA58453 standard; cDNA; 2958 BP.
XX
AC AAA58453;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human prostate cancer predisposing gene HPC2 coding sequence.
XX
KW Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW gene therapy; peptide therapy; drug design; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 51..2531
FT /*tag= a
FT /product= "HPC2"
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
DR P-PSDB; AAB07228.

XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Claim 3; Page 98-100; 157pp; English.
XX
CC The present sequence is the coding sequence of the human prostate
CC cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC alleles of this gene cause a predisposition to cancer, particularly
CC prostate cancer. This gene and its protein can be used in peptide and
CC gene therapy for cancer patients, as well as being useful as diagnostic
CC tools (both for cancer sufferers and those with a predisposition to the
CC disease) and in the production of cancer drugs. This sequence was
CC isolated by cloning and sequencing the region of the genome which
CC appeared to cause a predisposition to prostate cancer.
XX
SQ Sequence 2958 BP; 707 A; 805 C; 848 G; 598 T; 0 other;

Query Match 100.0%; Score 2481; DB 21; Length 2958;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     51 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 110

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    111 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 170

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    171 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 230

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    231 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 290

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    291 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 350

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    351 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 410

Qy    361 ttaagtggaatgattcttacttttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    411 ttaagtggaatgattcttacttttaaaggaaaccgggcttccaaagtgtgtactttctgga 470

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    471 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 530

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    531 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 590

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    591 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 650

Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    651 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 710

Qy    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    711 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 770

Qy    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    771 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 830

Qy    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    831 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 890

Qy    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    891 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 950

Qy    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    951 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 1010

Qy    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1011 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1070

Qy   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1071 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1130

Qy   1081 tttgggcctgacacccagcacttggtcctgaatgagaaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1131 tttgggcctgacacccagcacttggtcctgaatgagaaactgtgcctcagttcacaacctt 1190

Qy   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgctc 1250

Qy   1201 accagtttccgctgtaagaaggagggccccaccctcagtgtgcccatggttcagggtgaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1251 accagtttccgctgtaagaaggagggccccaccctcagtgtgcccatggttcagggtgaa 1310

Qy   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1311 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1370

Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1430

Qy   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1431 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1490

Qy   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1491 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1550

Qy   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1551 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1610

Qy   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1611 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1670

Qy   1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1671 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1730

Qy   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccctttgctggtggttgcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1731 cagagagaacgcgccttggcatctttgggaaagccgcttcacccctttgctggtggttgcc 1790

Qy   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1791 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1850

Qy   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1851 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1910
```

```
Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1911 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1970

Qy   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1971 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 2030

Qy   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2031 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2090

Qy   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2091 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2150

Qy   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2151 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2210

Qy   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2211 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2270

Qy   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2271 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2330

Qy   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2331 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2390

Qy   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2391 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2450

Qy   2401 ggcggcctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggcc 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2451 ggcggcctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggcc 2510

Qy   2461 aagaaggtcagagcccagtga 2481
          ||||||||||||||||||||
Db   2511 aagaaggtcagagcccagtga 2531
```

```
RESULT   2
AAA52810
ID   AAA52810 standard; cDNA; 2478 BP.
XX
AC   AAA52810;
XX
DT   20-SEP-2000  (first entry)
XX
DE   Human sulphatase G cDNA.
XX
KW   Human sulphatase G; hSG; chromosome 17p11.2; gene therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2478
FT                   /*tag=  a
FT                   /partial
FT                   /product= "hSG"
XX
PN   WO200034327-A1.
XX
PD   15-JUN-2000.
XX
PF   09-DEC-1999;   99WO-AU01092.
XX
PR   09-DEC-1998;   98AU-0007624.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Hopwood JJ,  Litjens T,  Hu RL;
XX
DR   WPI; 2000-431273/37.
DR   P-PSDB; AAY99850.
XX
PT   Novel isolated DNA sequence which encodes human sulfatase G or its
PT   fragment useful in gene therapy for treating patients suffering from
PT   sulfatase deficiency  -
XX
PS   Claim 2; Page 29-30; 33pp; English.
XX
CC   The present sequence encodes human sulphatase G (hSG). hSG is
CC   not a member of the well-characterised CTPSR sulphatase family.
CC   It belongs to a family showing sequence similarity to a sulphatase
CC   from the marine bacterium Alteromonas carrageenovora. The hSG gene
CC   contains 23 exons and is located at chromosome 17p11.2. The present
CC   sequence is clone lambda29.1 from a human testes cDNA library. It was
CC   isolated using human EST sequences with sequence similarity to the
CC   non-CTPSR family as a probe to screen the library. The cDNA insert was
CC   subcloned and the DNA sequence of both strands was determined. The
CC   sequence may be used to treat a patient suffering from hSG deficiency
CC   by replacing, repairing, or compensating for a DNA sequence within that
CC   patient's genome.
XX
SQ   Sequence 2478 BP; 587 A; 686 C; 709 G; 496 T; 0 other;
```

```
  Query Match             91.7%;  Score 2274;  DB 21;  Length 2478;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2474;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360

Qy    361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtacttttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtacttttctgga 420

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
```

```
Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagttgaatgaaaat 660

Qy    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720

Qy    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780

Qy    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840

Qy    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900

Qy    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960

Qy    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1020

Qy   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080

Qy   1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140

Qy   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgctc 1200

Qy   1201 accagtttccgctgtaagaaggagggccccacccctcagtgtgcccatggttcagggtgaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 accagtttccgctgtaagaaggagggccccacccctcagtgtgcccatggttcagggtgaa 1260

Qy   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320

Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380

Qy   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440

Qy   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500

Qy   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcacg 1560

Qy   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620

Qy   1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 actgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680

Qy   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740

Qy   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800

Qy   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860

Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1861 gtggaaagattgatcagttccctgttgcgaacatgtgatttggaagagtttcagacctgt 1920

Qy   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980

Qy   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040

Qy   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100

Qy   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160

Qy   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220

Qy   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280

Qy   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340

Qy   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400

Qy   2401 ggcggcctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggcc 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 ggcggcctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggcc 2460

Qy   2461 aagaaggtcagagcccag 2478
          ||||||||||||||||||
Db   2461 aagaaggtcagagcccag 2478
```

```
RESULT   3
AAH14250
ID   AAH14250 standard; cDNA; 2976 BP.
XX
AC   AAH14250;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:11557.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
```

PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 11557; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2976 BP; 712 A; 807 C; 856 G; 601 T; 0 other;

Query Match 81.5%; Score 2022; DB 22; Length 2976;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2412; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

```
Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 134

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 194

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 254

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 314

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 374

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    375 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 434

Qy    361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 494

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 554

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    555 atagaactggctatgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 614

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    615 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 674

Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    675 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 734

Qy    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    735 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 794

Qy    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    795 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 854

Qy    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    855 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 914

Qy    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    915 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 974

Qy    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    975 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 1034

Qy    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgccccctggcccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1035 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgccccctggcccttggtg 1094

Qy   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1095 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1154

Qy   1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1155 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1214

Qy   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttcccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1215 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttcccctgctc 1274

Qy   1201 accagtttccgctgtaagaaggagggccccaccctcagtgtgcccatggttcagggtgaa 1260
          ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db   1275 accagtttccgctgt---aaggagggccccaccctcagtgtgcccatggttcagggtgaa 1331

Qy   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1332 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1391

Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1451
```

```
Qy   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1452 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1511

Qy   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1512 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1571

Qy   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1572 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1631

Qy   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1632 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1691

Qy   1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1692 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1751

Qy   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1752 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1811

Qy   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1812 cccaaccagctcaaagcctggctccagcagtactacaaccagtgccaggaggtcctgcac 1871

Qy   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1931

Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1991

Qy   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1992 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 2051

Qy   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2052 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2111

Qy   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2112 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2171

Qy   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2172 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2231

Qy   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2232 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2291

Qy   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   2292 gagaaagtgggagttgcctttgaccacatgaaggtctgccttggagactttccaacaatg 2351

Qy   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2352 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2411

Qy   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2412 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2471

Qy   2401 ggcggcctggaggatggg 2418
          ||||||||||||||||||
Db   2472 ggcggcctggaggatggg 2489
```

```
RESULT   4
AAC76445
ID   AAC76445 standard; cDNA; 2546 BP.
XX
AC   AAC76445;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2000 polynucleotide sequence SEQ ID NO:3999.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42236.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 3179-3180; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
```

```
SQ   Sequence 2546 BP; 652 A; 643 C; 686 G; 564 T; 1 other;

  Query Match            64.5%;  Score 1600;  DB 21;  Length 2546;
  Best Local Similarity  99.9%;  Pred. No. 0;
  Matches 1720;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy     762 cttggtgctcaaagcaaaggagatgggcctcccagttgggacagctgccatcgctcccat 821
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 cttggtgctcaaagcaaaggagatgggcctcccagttgggacagctgccatcgctcccat 419

Qy     822 cattgctgctgtcaaggacgggaaaagcatcactcatgaaggaagagagattttggctga 881
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     420 cattgctgctgtcaaggacgggaaaagcatcactcatgaaggaagagagattttggctga 479

Qy     882 agagctgtgtactcctccagatcctggtgctgcttttgtggtggtagaatgtccagatga 941
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 agagctgtgtactcctccagatcctggtgctgcttttgtggtggtagaatgtccagatga 539

Qy     942 aagcttcattcaacccatctgtgagaatgccacctttcagaggtaccaaggaaaggcaga 1001
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     540 aagcttcattcaacccatctgtgagaatgccacctttcagaggtaccaaggaaaggcaga 599

Qy    1002 tgcccccgtggccttggtggttcacatggccccagcatctgtgcttgtggacagcaggta 1061
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 tgcccccgtggccttggtggttcacatggccccagcatctgtgcttgtggacagcaggta 659

Qy    1062 ccagcagtggatggagaggtttgggcctgacacccagcacttggtcctgaatgagaactg 1121
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     660 ccagcagtggatggagaggtttgggcctgacacccagcacttggtcctgaatgagaactg 719

Qy    1122 tgcctcagttcacaaccttcgcagccacaagattcaaacccagctcaacctcatccaccc 1181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     720 tgcctcagttcacaaccttcgcagccacaagattcaaacccagctcaacctcatccaccc 779

Qy    1182 ggacatcttccccctgctcaccagtttccgctgtaagaaggagggccccaccctcagtgt 1241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     780 ggacatcttccccctgctcaccagtttccgctgtaagaaggagggccccaccctcagtgt 839

Qy    1242 gcccatggttcagggtgaatgcctcctcaagtaccagctccgtcccaggagggagtggca 1301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     840 gcccatggttcagggtgaatgcctcctcaagtaccagctccgtcccaggagggagtggca 899

Qy    1302 gagggatgccattattacttgcaatcctgaggaattcatagttgaggcgctgcagcttcc 1361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     900 gagggatgccattattacttgcaatcctgaggaattcatagttgaggcgctgcagcttcc 959

Qy    1362 caacttccagcagagcgtgcaggagtacaggaggagtgcgcaggacggcccagccccagc 1421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     960 caacttccagcagagcgtgcaggagtacaggaggagtgcgcaggacggcccagccccagc 1019

Qy    1422 agagaaaagaagtcagtacccagaaatcatcttccttggaacagggtctgccatcccgat 1481
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1020 agagaaaagaagtcagtacccagaaatcatcttccttggaacagggtctgccatcccgat 1079

Qy    1482 gaagattcgaaatgtcagtgccacacttgtcaacataagccccgacacgtctctgctact 1541
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1080 gaagattcgaaatgtcagtgccacacttgtcaacataagccccgacacgtctctgctact 1139

Qy    1542 ggactgtggtgagggcacatttgggcagctgtgccgtcattacggagaccaggtggacag 1601
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1140 ggactgtggtgagggcacatttgggcagctgtgccgtcattacggagaccaggtggacag 1199

Qy    1602 ggtcctgggcaccctggctgctgtgtttgtgtcccacctgcacgcagatcaccacacggg 1661
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1200 ggtcctgggcaccctggctgctgtgtttgtgtcccacctgcacgcagatcaccacacggg 1259

Qy    1662 cttgccaagtatcttgctgcagagagaacgcgccttggcatctttgggaaagccgcttca 1721
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1260 cttgccaagtatcttgctgcagagagaacgcgccttggcatctttgggaaagccgcttca 1319

Qy    1722 ccctttgctggtggttgcccccaaccagctcaaagcctggctccagcagtaccacaacca 1781
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    1320 ccctttgctggtggttgcccccaaccagctcaaagcctggctccagcagtaccacaacca 1379

Qy    1782 gtgccaggaggtcctgcaccacatcagtatgattcctgccaaatgccttcaggaaggggc 1841
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1380 gtgccaggaggtcctgcaccacatcagtatgattcctgccaaatgccttcaggaaggggc 1439

Qy    1842 tgagatctccagtcctgcagtggaaagattgatcagttcgctgttgcgaacatgtgattt 1901
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1440 tgagatctccagtcctgcagtggaaagattgatcagttcgctgttgcgaacatgtgattt 1499

Qy    1902 ggaagagtttcagacctgtctggtgcggcactgcaagcatgcgtttggctgtgcgctggt 1961
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1500 ggaagagtttcagacctgtctggtgcggcactgcaagcatgcgtttggctgtgcgctggt 1559

Qy    1962 gcacacctctggctggaaagtggtctattccggggacaccatgccctgcgaggctctggt 2021
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1560 gcacacctctggctggaaagtggtctattccggggacaccatgccctgcgaggctctggt 1619

Qy    2022 ccggatggggaaagatgccaccctcctgatacatgaagccaccctggaagatggtttgga 2081
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1620 ccggatggggaaagatgccaccctcctgatacatgaagccaccctggaagatggtttgga 1679

Qy    2082 agaggaagcagtggaaaagacacacagcacaacgtcccaagccatcagcgtggggatgcg 2141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1680 agaggaagcagtggaaaagacacacagcacaacgtcccaagccatcagcgtggggatgcg 1739

Qy    2142 gatgaacgcggagttcattatgctgaaccacttcagccagcgctatgccaaggtccccct 2201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1740 gatgaacgcggagttcattatgctgaaccacttcagccagcgctatgccaaggtccccct 1799

Qy    2202 cttcagccccaacttcagcgagaaagtgggagttgcctttgaccacatgaaggtctgctt 2261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1800 cttcagccccaacttcagcgagaaagtgggagttgcctttgaccacatgaaggtctgctt 1859

Qy    2262 tggagactttccaacaatgcccaagctgatt-cccccactgaaagccctgtttgctggcg 2320
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    1860 tggagactttccaacaatgcccaagctgattcccccccactgaaagccctgtttgctggcg 1919

Qy    2321 acatcgaggagatggaggagcgcagggagaagcgggagctgcggcaggtgcgggcggccc 2380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1920 acatcgaggagatggaggagcgcagggagaagcgggagctgcggcaggtgcgggcggccc 1979

Qy    2381 tcctgtccagggagctggcaggcggcctggaggatggggagcctcagcagaagcgggccc 2440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1980 tcctgtccagggagctggcaggcggcctggaggatggggagcctcagcagaagcgggccc 2039

Qy    2441 acacagaggagccacaggccaagaaggtcagagcccagtga 2481
           |||||||||||||||||||||||||||||||||||||||||
Db    2040 acacagaggagccacaggccaagaaggtcagagcccagtga 2080


RESULT   5
AAH05835
ID   AAH05835 standard; cDNA; 584 BP.
XX
AC   AAH05835;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:2670.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
```

```
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 2670; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 584 BP; 122 A; 166 C; 178 G; 115 T; 3 other;


  Query Match             14.7%;  Score 364;  DB 22;  Length 584;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-168;
  Matches  364;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 134

Qy      61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     135 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 194

Qy     121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     195 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 254

Qy     181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     255 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 314

Qy     241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 374

Qy     301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     375 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 434

Qy     361 ttaa 364
           ||||
Db     435 ttaa 438


RESULT   6
AAA60207
ID   AAA60207 standard; DNA; 26664 BP.
XX
AC   AAA60207;
XX
DT   07-DEC-2000  (first entry)
XX
DE   Human prostate cancer predisposing gene HPC2 genomic sequence.
XX
KW   Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW   gene therapy; peptide therapy; drug design; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             910..26039
FT                   /*tag= a
FT                   /product= "HPC2"
FT                   /note= "this sequence contains introns"
FT                   /transl_except= (pos:23892..23895,aa:Glu)
FT   exon            910..1154
FT                   /*tag= b
FT                   /number= 1
FT   exon            1736..1786
FT                   /*tag= c
FT                   /number= 2
FT   exon            1925..1995
FT                   /*tag= d
FT                   /number= 3
FT   exon            3025..3089
FT                   /*tag= e
FT                   /number= 4
FT   exon            4361..4418
FT                   /*tag= f
FT                   /number= 5
FT   exon            5582..5650
FT                   /*tag= g
FT                   /number= 6
FT   exon            7075..7194
FT                   /*tag= h
FT                   /number= 7
FT   exon            8186..8244
FT                   /*tag= i
FT                   /number= 8
FT   exon            12878..12936
FT                   /*tag= j
FT                   /number= 9
FT   exon            13032..13104
FT                   /*tag= k
FT                   /number= 10
FT   exon            13756..13868
FT                   /*tag= l
FT                   /number= 11
FT   exon            15283..15378
FT                   /*tag= m
FT                   /number= 12
FT   exon            16278..16416
FT                   /*tag= n
FT                   /number= 13
FT   exon            16498..16583
FT                   /*tag= o
FT                   /number= 14
FT   exon            18583..18701
FT                   /*tag= p
FT                   /number= 15
FT   exon            20349..20445
FT                   /*tag= q
FT                   /number= 16
```

```
FT   exon            22172..22310
FT                   /*tag= r
FT                   /number= 17
FT   exon            22879..22917
FT                   /*tag= s
FT                   /number= 18
FT   exon            23045..23154
FT                   /*tag= t
FT                   /number= 19
FT   exon            23795..23895
FT                   /*tag= u
FT                   /number= 20
FT   exon            23973..24093
FT                   /*tag= v
FT                   /number= 21
FT   exon            24354..24432
FT                   /*tag= w
FT                   /number= 22
FT   exon            25026..25170
FT                   /*tag= x
FT                   /number= 23
FT   exon            25812..26036
FT                   /*tag= y
FT                   /number= 24
FT   polyA_signal    26447..26452
FT                   /*tag= z
XX
PN   WO200027864-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26055.
XX
PR   06-NOV-1998;   98US-0107468.
XX
PA   (MYRI-) MYRIAD GENETICS INC.
XX
PI   Tavtigian SV,  Teng DHF,  Simard J,  Rommens JM;
XX
DR   WPI; 2000-376481/32.
DR   P-PSDB; AAB07228.
XX
PT   Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT   antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS   Claim 3; Page 108-122; 157pp; English.
XX
CC   The present sequence is the genomic sequence of the human prostate
CC   cancer predisposing gene HPC2, which is found on chromosome 17p. Some
CC   alleles of this gene cause a predisposition to cancer, particularly
CC   prostate cancer. This gene and its protein can be used in peptide and
CC   gene therapy for cancer patients, as well as being useful as diagnostic
CC   tools (both for cancer sufferers and those with a predisposition to the
CC   disease) and in the production of cancer drugs. This sequence was
CC   isolated by cloning and sequencing the region of the genome which
CC   appeared to cause a predisposition to prostate cancer.
XX
SQ   Sequence 26664 BP; 6173 A; 6300 C; 6519 G; 7661 T; 11 other;

  Query Match             10.0%;  Score 247;  DB 21;  Length 26664;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-111;
  Matches  247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     910 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 969

Qy      61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     970 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 1029

Qy     121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1030 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 1089

Qy     181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1090 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 1149

Qy     241 aaccggt 247
           |||||||
Db    1150 aaccggt 1156


RESULT   7
AAZ80231
ID   AAZ80231 standard; cDNA; 238 BP.
XX
AC   AAZ80231;
XX
DT   07-APR-2000  (first entry)
XX
DE   Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
XX
KW   Human; gene expression product; diagnosis; tumour; colon cancer;
KW   colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW   cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW   hyperplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO9964576-A2.
XX
PD   16-DEC-1999.
XX
PF   09-JUN-1999;   99WO-IB01062.
XX
PR   10-JUN-1998;   98US-0088801.
XX
PA   (FARB ) BAYER CORP.
XX
PI   Endege WO,  Steinmann KE,  Astle JH,  Burgess CC,  Bushnell SE;
PI   Carroll E,  Catino TJ,  Derti A,  Ford DM,  Lewis ME,  Monahan JE;
PI   Schlegel R;
XX
DR   WPI; 2000-087220/07.
XX
PT   Novel nucleic acids, used to develop products for the diagnosis and
PT   treatment of disorders involving unwanted cell proliferation,
PT   particularly cancers, especially colon cancer -
XX
PS   Claim 15; Page 258; 469pp; English.
XX
CC   AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from
CC   the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC   cDNA clones can be used to generate antisense oligonucleotides which
CC   can be used for antisense therapy. Methods and products from the present
CC   invention can be used for identifying and/or classifying cancerous cells
CC   present in a human tumour, particularly in solid tumours, e.g.
CC   carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones
CC   can be used for developing agents for the diagnosis and treatment of
CC   disorders involving unwanted cell proliferation, such as neoplasia,
CC   dysplasia or hyperplasia.
XX
SQ   Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;

  Query Match              9.6%;  Score 237;  DB 21;  Length 238;
  Best Local Similarity   100.0%;  Pred. No. 6.6e-106;
  Matches  237;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     176 acctgcaggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccg 235
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 acctgcaggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccg 60
```

```
Qy    236 agttcaaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcaca 295
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 agttcaaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcaca 120

Qy    296 agttaaaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttg 355
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 agttaaaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttg 180

Qy    356 ggggcttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtac 412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ggggcttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtac 237
```

RESULT 8
AAA60390
ID AAA60390 standard; cDNA; 326 BP.
XX
AC AAA60390;
XX
DT 07-DEC-2000 (first entry)
XX
DE Murine prostate cancer predisposing gene HPC2 coding sequence.
XX
KW Mouse; prostate cancer predisposing gene; HPC2; gene therapy;
KW peptide therapy; drug design; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 51..269
FT /*tag= a
FT /product= "Mm.HPC2"
FT /partial
FT /note= "this sequence contains no termination codon"
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
DR P-PSDB; AAB07230.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Disclosure; Page 151; 157pp; English.
XX
CC The present sequence is the coding sequence of the murine prostate
CC cancer predisposing gene Mm. HPC2, the human homologue of which is found
CC on chromosome 17p. Some alleles of this gene cause a predisposition to
CC cancer, particularly prostate cancer. This gene and its protein can be
CC used in peptide and gene therapy for cancer patients, as well as being
CC useful as diagnostic tools (both for cancer sufferers and those with a
CC predisposition to the disease) and in the production of cancer drugs.
XX
SQ Sequence 326 BP; 39 A; 104 C; 127 G; 56 T; 0 other;

Query Match 1.0%; Score 26; DB 21; Length 326;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    166 aacaccgtgtacctgcaggtggtggc 191
          ||||||||||||||||||||||||||
Db    192 aacaccgtgtacctgcaggtggtggc 217
```

RESULT 9
AAA60373
ID AAA60373 standard; cDNA; 23 BP.
XX
AC AAA60373;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human prostate cancer predisposing gene HPC2 variant 1641insG (normal).
XX
KW Human; prostate cancer predisposing gene; HPC2; chromosome 17p;
KW gene therapy; peptide therapy; drug design; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Example 5; Page 63; 157pp; English.
XX
CC The present sequence is part of the coding sequence of a variant of the
CC human prostate cancer predisposing gene HPC2, which is found on
CC chromosome 17p. This variant has been designated 1641insG. Some
CC alleles of this gene cause a predisposition to cancer, particularly
CC prostate cancer, this allele being an example of this. The HPC2 gene and
CC its protein can be used in peptide and gene therapy for cancer patients,
CC as well as being useful as diagnostic tools (both for cancer sufferers
CC and those with a predisposition to the disease) and in the production of
CC cancer drugs. This sequence was isolated by mutation screening of the
CC HPC2 gene in humans.
XX
SQ Sequence 23 BP; 5 A; 9 C; 5 G; 4 T; 0 other;

Query Match 0.9%; Score 23; DB 21; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.43;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1630 gtgtcccacctgcacgcagatca 1652
          |||||||||||||||||||||||
Db      1 gtgtcccacctgcacgcagatca 23
```

RESULT 10
AAC24564/c
ID AAC24564 standard; cDNA; 143 BP.
XX
AC AAC24564;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 28639.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.

XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 28639; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 143 BP; 40 A; 26 C; 45 G; 32 T; 0 other;

Query Match 0.8%; Score 21; DB 21; Length 143;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 976 tttcagaggtaccaaggaaag 996
||||||||||||||||||||
Db 67 TTTCAGAGGTACCAAGGAAAG 47

RESULT 11
AAA60275/c
ID AAA60275 standard; DNA; 38 BP.
XX
AC AAA60275;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA exon 1 mutation screening primer SEQ ID NO: 96.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Example 5; Page 59; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 38 BP; 10 A; 9 C; 11 G; 8 T; 0 other;

Query Match 0.8%; Score 20; DB 21; Length 38;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 gacgcaccatatcgcaggca 75
||||||||||||||||||||
Db 38 GACGCACCATATCGCAGGCA 19

RESULT 12
AAF09297
ID AAF09297 standard; cDNA; 308 BP.
XX
AC AAF09297;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:1820.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 86; Page 1064; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal

CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discqvered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 308 BP; 88 A; 56 C; 100 G; 63 T; 1 other;

Query Match 0.8%; Score 20; DB 21; Length 308;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2334 ggaggagcgcagggagaagc 2353
||||||||||||||||||||
Db 30 ggaggagcgcagggagaagc 49

RESULT 13
AAC57035
ID AAC57035 standard; DNA; 457 BP.
XX
AC AAC57035;
XX
DT 25-JAN-2001 (first entry)
XX
DE Pinus radiata transcription factor DNA sequence #481.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Pinus radiata.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
PS Claim 1; Pages 570-571; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 457 BP; 130 A; 94 C; 118 G; 115 T; 0 other;

Query Match 0.8%; Score 20; DB 21; Length 457;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2082 agaggaagcagtggaaaaga 2101
||||||||||||||||||||
Db 146 agaggaagcagtggaaaaga 165

RESULT 14
AAA60211/c
ID AAA60211 standard; DNA; 19 BP.
XX
AC AAA60211;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA 5' RACE primer SEQ ID NO: 32.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Example 3; Page 55; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 19 BP; 4 A; 7 C; 3 G; 5 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    257 actgtggagaaggcgttca 275
          ||||||||||||||||||
Db     19 ACTGTGGAGAAGGCGTTCA 1
```

RESULT 15
AAA60229
ID AAA60229 standard; DNA; 19 BP.
XX
AC AAA60229;
XX
DT 07-DEC-2000 (first entry)
XX
DE Human HPC2 cDNA sequencing primer SEQ ID NO: 50.
XX
KW Human; mouse; prostate cancer predisposing gene; HPC2;
KW human chromosome 17p; gene therapy; peptide therapy; drug design;
KW PCR primer; sequencing primer; ss.
XX
OS Homo sapiens.
XX
PN WO200027864-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26055.
XX
PR 06-NOV-1998; 98US-0107468.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI Tavtigian SV, Teng DHF, Simard J, Rommens JM;
XX
DR WPI; 2000-376481/32.
XX
PT Human prostate cancer (HPC)2 nucleic acids, polypeptides, and
PT antibodies, useful for treatment and diagnosis of prostate cancer -
XX
PS Example 3; Page 55; 157pp; English.
XX
CC The present sequence is a primer used in the isolation of the human
CC and murine prostate cancer predisposing genes HPC2 and Mm.HPC2. The human
CC version of the gene is found on chromosome 17p. Some alleles cause a
CC predisposition to cancer, particularly prostate cancer. This gene and its
CC protein can be used in peptide and gene therapy for cancer patients, as
CC well as being useful as diagnostic tools (both for cancer sufferers and
CC those with a predisposition to the disease) and in the production of
CC cancer drugs.
XX
SQ Sequence 19 BP; 5 A; 3 C; 6 G; 5 T; 0 other;

Query Match 0.8%; Score 19; DB 21; Length 19;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1901 tggaagagtttcagacctg 1919
          ||||||||||||||||||
Db      1 tggaagagtttcagacctg 19
```

Search completed: April 1, 2002, 11:11:03
Job time: 5877 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: April 1, 2002, 09:32:36 ; Search time 1802.89 Seconds
(without alignments)
22702.163 Million cell updates/sec

Title: US-09-988-687-1
Perfect score: 2481
Sequence: 1 atgtgggcgctttgctcgct.....agaaggtcagagcccagtga 2481

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vi:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description
	1	2481	100.0	2908	9	AF304370	AF304370 Homo sapi
	2	2379	95.9	2997	9	BC001939	BC001939 Homo sapi
	3	2379	95.9	3006	9	BC004158	BC004158 Homo sapi
	4	2022	81.5	2976	9	AK001392	AK001392 Homo sapi
	5	1667	67.2	2908	9	AF308698	AF308698 Pan trogl
	6	1266	51.0	2893	9	AF308694	AF308694 Gorilla g
	7	247	10.0	118788	9	AC005277	AC005277 Homo sapi
	8	142	5.7	740	9	AF304371S2	AF304369 Homo sapi
	9	122	4.9	721	9	AF304371S1	AF304371 Homo sapi
	10	47	1.9	2712	10	AF308696	AF308696 Mus muscu
	11	46	1.9	34593	10	AF348157	AF348157 Mus muscu
c	12	23	0.9	43514	9	AC005545	AC005545 Homo sapi
	13	23	0.9	105490	2	AF165142	AF165142 Homo sapi
	14	23	0.9	190195	2	AC091560	AC091560 Homo sapi
	15	23	0.9	190705	2	AC019331	AC019331 Homo sapi
c	16	21	0.8	64151	9	AL139399	AL139399 Human DNA
	17	21	0.8	125403	9	AC008929	AC008929 Homo sapi
c	18	21	0.8	148193	9	AC027347	AC027347 Homo sapi
	19	21	0.8	165665	2	AL442223	AL442223 Homo sapi
	20	21	0.8	196424	9	AP000719	AP000719 Homo sapi
c	21	21	0.8	196613	2	AP002490	AP002490 Homo sapi
	22	21	0.8	220146	2	AC034179	AC034179 Homo sapi
	23	20	0.8	1176	5	AB046207	AB046207 Anguilla
	24	20	0.8	1573	8	ATU23796	U23796 Arabidopsis
	25	20	0.8	4918	8	AF047031	AF047031 Arabidops
	26	20	0.8	6718	9	AB014516	AB014516 Homo sapi
c	27	20	0.8	13800	1	SC7B7	AL009199 Streptomy
	28	20	0.8	46275	9	AC003107	AC003107 Human DNA
c	29	20	0.8	76656	9	AL441923	AL441923 Human DNA
	30	20	0.8	79590	8	AB007727	AB007727 Arabidops
	31	20	0.8	84129	8	AB013392	AB013392 Arabidops
	32	20	0.8	103638	8	ATF14F18	AL163812 Arabidops
	33	20	0.8	114521	9	AL159985	AL159985 Human DNA
c	34	20	0.8	117338	9	HS173D1	AL031984 Human DNA
	35	20	0.8	127953	2	AC022835	AC022835 Homo sapi
c	36	20	0.8	148508	2	AC027795	AC027795 Homo sapi
	37	20	0.8	149006	9	AC025812	AC025812 Homo sapi
	38	20	0.8	149180	2	AC015525	AC015525 Homo sapi
c	39	20	0.8	150683	2	AC069420	AC069420 Homo sapi
	40	20	0.8	165490	2	AC025326	AC025326 Homo sapi
c	41	20	0.8	165493	2	AC018876	AC018876 Homo sapi
c	42	20	0.8	166304	2	AC091135	AC091135 Homo sapi
c	43	20	0.8	166344	2	AL162577	AL162577 Homo sapi
	44	20	0.8	168997	2	AC021927	AC021927 Homo sapi
	45	20	0.8	177868	2	AC025191	AC025191 Homo sapi

ALIGNMENTS

RESULT 1
AF304370
LOCUS AF304370 2908 bp mRNA PRI 23-FEB-2001
DEFINITION Homo sapiens putative prostate cancer susceptibility protein HPC2/ELAC2 mRNA, complete cds.
ACCESSION AF304370
VERSION AF304370.1 GI:10880932
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2908)
AUTHORS Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,

Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE A candidate prostate cancer susceptibility gene at chromosome 17p
JOURNAL Nat. Genet. 27 (2), 172-180 (2001)
PUBMED 11175785
REFERENCE 2 (bases 1 to 2908)
AUTHORS Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
TITLE Direct Submission
JOURNAL Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt Lake City, UT 84108, USA
FEATURES Location/Qualifiers
source 1. .2908
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p"
CDS 1. .2481
/codon_start=1
/product="putative prostate cancer susceptibility protein HPC2/ELAC2"
/protein_id="AAG24441.1"
/db_xref="GI:10880933"
/translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESN
ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
ASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT 702 a 793 c 826 g 587 t
ORIGIN

Query Match 100.0%; Score 2481; DB 9; Length 2908;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 60

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 120

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 180

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 240

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 300

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 360

Qy    361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 420

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 480

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 540

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 600

Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTCGAATGAAAAT 660

Qy    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGACTCTTCCCTGGTC 720

Qy    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGCTCAAAGCAAAG 780

Qy    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGAC 840

Qy    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGGAAAAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 900

Qy    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATC 960

Qy    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgccccctgtggccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTG 1020

Qy   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGG 1080

Qy   1081 tttgggcctgacacccagcacttggtcctgaatgagaaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTT 1140

Qy   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1200

Qy   1201 accagtttccgctgtaagaaggagggccccacccctcagtgtgcccatggttcagggtgaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAA 1260

Qy   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACT 1320

Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 TGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTG 1380

Qy   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
```

```
Db   1381 CAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTAC 1440

Qy   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1500

Qy   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACA 1560

Qy   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCT 1620

Qy   1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTG 1680

Qy   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCC 1740

Qy   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1800

Qy   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCA 1860

Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1920

Qy   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAA 1980

Qy   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCC 2040

Qy   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 ACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2100

Qy   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATT 2160

Qy   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGC 2220

Qy   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGAGACTTTCCAACAATG 2280

Qy   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CCCAAGCTGATTCCCCCACTGAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAG 2340

Qy   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGCAGGGAGAAGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2400

Qy   2401 ggcggcctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggcc 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGAGCCACAGGCC 2460

Qy   2461 aagaaggtcagagcccagtga 2481
          ||||||||||||||||||||
```

```
Db   2461 AAGAAGGTCAGAGCCCAGTGA 2481


RESULT   2
BC001939
LOCUS       BC001939     2997 bp    mRNA            PRI       12-JUL-2001
DEFINITION  Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:4102 IMAGE:2820640, mRNA, complete cds.
ACCESSION   BC001939
VERSION     BC001939.1  GI:12804972
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2997)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 9 Row: i Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7022621.
FEATURES             Location/Qualifiers
     source          1. .2997
                     /organism="Homo sapiens"
                     /db_xref="LocusID:60528"
                     /db_xref="taxon:9606"
                     /clone="MGC:4102 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             61. .2541
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH01939.1"
                     /db_xref="GI:12804973"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSELN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
```

EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT 740 a 806 c 853 g 598 t
ORIGIN

Query Match 95.9%; Score 2379; DB 9; Length 2997;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2479; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 120

Qy      61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 180

Qy     121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 240

Qy     181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 300

Qy     241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 360

Qy     301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 420

Qy     361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 480

Qy     421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 540

Qy     481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 600

Qy     541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 660

Qy     601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     661 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTTGAATGAAAAT 720

Qy     661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGACTCTTCCCTGGTC 780

Qy     721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGCTCAAAGCAAAG 840

Qy     781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGAC 900

Qy     841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 GGGAAAAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 960

Qy     901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATC 1020

Qy     961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 TGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTG 1080

Qy    1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGG 1140

Qy    1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTT 1200

Qy    1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1260

Qy    1201 accagtttccgctgtaagaaggagggccccaccctcagtgtgcccatggttcagggtgaa 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAA 1320

Qy    1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACT 1380

Qy    1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTG 1440

Qy    1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTAC 1500

Qy    1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1560

Qy    1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACG 1620

Qy    1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 TTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCT 1680

Qy    1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTG 1740

Qy    1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCC 1800

Qy    1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 CCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1860

Qy    1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCA 1920

Qy    1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1980

Qy    1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 CTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAA 2040

Qy    1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 GTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCC 2100
```

```
Qy   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2160

Qy   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATT 2220

Qy   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGC 2280

Qy   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 GAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGAGACTTTCCAACAATG 2340

Qy   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CCCAAGCTGATTCCCCCACTGAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAG 2400

Qy   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CGCAGGGAGAAGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2460

Qy   2401 ggcggcctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggcc 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGAGCCACAGGCC 2520

Qy   2461 aagaaggtcagagcccagtga 2481
          ||||||||||||||||||||
Db   2521 AAGAAGGTCAGAGCCCAGTGA 2541
```

```
RESULT   3
BC004158
LOCUS       BC004158     3006 bp    mRNA            PRI       12-JUL-2001
DEFINITION  Homo sapiens, putative prostate cancer susceptibility protein,
            clone MGC:2441 IMAGE:2820640, mRNA, complete cds.
ACCESSION   BC004158
VERSION     BC004158.1  GI:13278770
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3006)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 2 Row: c Column: 13
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 7022621.
FEATURES             Location/Qualifiers
     source          1. .3006
                     /organism="Homo sapiens"
                     /db_xref="LocusID:60528"
                     /db_xref="taxon:9606"
                     /clone="MGC:2441 IMAGE:2820640"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             70. .2550
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAH04158.1"
                     /db_xref="GI:13278771"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSELN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFRCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERAL
                     ASLGKPLHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFS
                     EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
                     LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT      742 a    808 c    858 g    598 t
ORIGIN
```

```
  Query Match             95.9%;  Score 2379;  DB 9;  Length 3006;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2479;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 129

Qy      61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     130 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 189

Qy     121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     190 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 249

Qy     181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     250 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 309

Qy     241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     310 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 369

Qy     301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     370 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 429

Qy     361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     430 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 489

Qy     421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     490 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 549

Qy     481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     550 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 609

Qy     541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     610 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 669
```

```
Qy   601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   670 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTTGAATGAAAAT 729

Qy   661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   730 GAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGACTCTTCCCTGGTC 789

Qy   721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   790 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGCTCAAAGCAAAG 849

Qy   781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   850 GAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGAC 909

Qy   841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   910 GGGAAAAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 969

Qy   901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   970 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATC 1029

Qy   961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1030 TGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTG 1089

Qy  1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1090 GTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGG 1149

Qy  1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1150 TTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTT 1209

Qy  1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttccccctgctc 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1210 CGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1269

Qy  1201 accagtttccgctgtaagaaggagggcccccaccctcagtgtgcccatggttcagggtgaa 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1270 ACCAGTTTCCGCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAA 1329

Qy  1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1330 TGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACT 1389

Qy  1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1390 TGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTG 1449

Qy  1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1450 CAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTAC 1509

Qy  1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1510 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1569

Qy  1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1570 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACG 1629

Qy  1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1630 TTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCT 1689

Qy  1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1690 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTG 1749

Qy  1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1750 CAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCC 1809

Qy  1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1810 CCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1869

Qy  1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1870 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCA 1929

Qy  1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1930 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1989

Qy  1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1990 CTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAA 2049

Qy  1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2050 GTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCC 2109

Qy  2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2110 ACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2169

Qy  2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2170 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATT 2229

Qy  2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2230 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGC 2289

Qy  2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2290 GAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGAGACTTTCCAACAATG 2349

Qy  2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2350 CCCAAGCTGATTCCCCCACTGAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAG 2409

Qy  2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2410 CGCAGGGAGAAGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2469

Qy  2401 ggcggcctggaggatggggagcctcagcagaagcgggcccacacagaggagccacaggcc 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2470 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAAGCGGGCCCACACAGAGGAGCCACAGGCC 2529

Qy  2461 aagaaggtcagagcccagtga 2481
         ||||||||||||||||||||
Db  2530 AAGAAGGTCAGAGCCCAGTGA 2550
```

```
RESULT   4
AK001392
LOCUS       AK001392     2976 bp    mRNA            PRI       22-FEB-2000
DEFINITION  Homo sapiens cDNA FLJ10530 fis, clone NT2RP2000985.
ACCESSION   AK001392
VERSION     AK001392.1  GI:7022621
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
            clone_lib:NT2RP2 clone:NT2RP2000985.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
            Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
            Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
```

```
            Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
            Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
            Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
            Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished (2000)
REFERENCE   2  (bases 1 to 2976)
  AUTHORS   Isogai,T. and Otsuki,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) to the DDBJ/EMBL/GenBank databases. Takao
            Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,
            Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
            Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing and clone selection:
            Helix Research Institute (supported by Japan Key Technology Center
            etc.) and Department of Virology, Institute of Medical Science,
            University of Tokyo.
FEATURES             Location/Qualifiers
     source          1. .2976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone="NT2RP2000985"
                     /clone_lib="NT2RP2"
                     /note="cloning vector: pME18SFL3; mRNA from NT2 neuronal
                     precursor cells after 2-weeks retinoic acid (RA)
                     induction."
     CDS             75. .2639
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAA91666.1"
                     /db_xref="GI:7022622"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAMRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPASVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFRCKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEAL
                     QLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISPD
                     TSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLPSILLQRERALA
                     SLGKPLHPLLVVAPNQLKAWLQQYYNQCQEVLHHISMIPAKCLQEGAEISSPAVERLI
                     SSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATLL
                     IHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFSE
                     KVGVAFDHMKVCLGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSREL
                     AGGLEDGSLSRSGPTQRSHRPRRSEPSEDLGDPELRRLCVFCPTHAPVSALLAGRS"
BASE COUNT      712 a    807 c    856 g    601 t
ORIGIN


  Query Match             81.5%;  Score 2022;  DB 9;  Length 2976;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2412;  Conservative    0;  Mismatches    3;  Indels    3;  Gaps    1;

Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 134

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 194

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 254

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db    255 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 314

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 374

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    375 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 434

Qy    361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 494

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 554

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    555 ATAGAACTGGCTATGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 614

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    615 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 674

Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    675 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTCGAATGAAAAT 734

Qy    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    735 GAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGACTCTTCCCTGGTC 794

Qy    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    795 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGCTCAAAGCAAAG 854

Qy    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    855 GAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGAC 914

Qy    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    915 GGGAAAAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 974

Qy    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    975 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATC 1034

Qy    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1035 TGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTG 1094

Qy   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1095 GTTCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGG 1154

Qy   1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1155 TTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTT 1214

Qy   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttcccccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1215 CGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1274

Qy   1201 accagtttccgctgtaagaaggagggccccacctcagtgtgcccatggttcagggtgaa 1260
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1275 ACCAGTTTCCGCTGT---AAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAA 1331

Qy   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1332 TGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACT 1391
```

```
Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 TGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTG 1451

Qy   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1452 CAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTAC 1511

Qy   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1512 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1571

Qy   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1572 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACA 1631

Qy   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1632 TTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCT 1691

Qy   1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1692 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCCAAGTATCTTGCTG 1751

Qy   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1752 CAGAGAGAACGCGCCTTGGCATCTTTGGGAAAGCCGCTTCACCCTTTGCTGGTGGTTGCC 1811

Qy   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1812 CCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACTACAACCAGTGCCAGGAGGTCCTGCAC 1871

Qy   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCA 1931

Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1991

Qy   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1992 CTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAA 2051

Qy   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2052 GTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCC 2111

Qy   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2112 ACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2171

Qy   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2172 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATT 2231

Qy   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2232 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAGC 2291

Qy   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2292 GAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCCTTGGAGACTTTCCAACAATG 2351

Qy   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2352 CCCAAGCTGATTCCCCCACTGAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAG 2411

Qy   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2412 CGCAGGGAGAAGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2471
```

```
Qy   2401 ggcggcctggaggatggg 2418
          ||||||||||||||||||
Db   2472 GGCGGCCTGGAGGATGGG 2489
```

```
RESULT   5
AF308698
LOCUS       AF308698     2908 bp    mRNA            PRI       27-FEB-2001
DEFINITION  Pan troglodytes ELAC2 mRNA, complete cds.
ACCESSION   AF308698
VERSION     AF308698.1  GI:10946496
KEYWORDS    .
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Pan troglodytes ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2908)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES             Location/Qualifiers
     source          1. .2908
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     CDS             1. .2481
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24920.1"
                     /db_xref="GI:10946497"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGIQRLMQEHKLKV
                     ARLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGKHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPESVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFPCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIIEA
                     LQLPNFQQSVQEYRRSAQDGPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLLNILLQRERAL
                     ASLGKPFHPLLVVAPNQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
```

LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFN
EKVGVAFDHMKVCFGDFATMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSRE
LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT 712 a 788 c 819 g 589 t
ORIGIN

Query Match 67.2%; Score 1667; DB 9; Length 2908;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2417; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```
Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 60

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 120

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 180

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 240

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    241 AACCGGTATCTCTTCAACTGTGGAGAAGGCATTCAGAGACTCATGCAGGAGCACAAGTTA 300

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGGTTGCTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 360

Qy    361 ttaagtggaatgattcttacttttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 420

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTCCACAACTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 480

Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 540

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCA 600

Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    601 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTCAAATGAAAAT 660

Qy    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGACTCTTCCCTGGTC 720

Qy    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGCTCAAAGCAAAG 780

Qy    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGAC 840

Qy    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGGAAAAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 900

Qy    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATC 960

Qy    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgccccctggccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTG 1020

Qy   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCACATGGCCCCAGAATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGG 1080

Qy   1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTT 1140

Qy   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttcccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1200

Qy   1201 accagtttccgctgtaagaaggagggccccaccctcagtgtgcccatggttcagggtgaa 1260
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAGTTTCCCCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAA 1260

Qy   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAGTGGCAGAGGGATGCCATTATTACT 1320

Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          |||||||||||||||||||| |||||||||||||||||||||||||||||||| |||||
Db   1321 TGCAATCCTGAGGAATTCATAATTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGTGTG 1380

Qy   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAAAGAAGTCAGTAC 1440

Qy   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCGATGAAGATTCGAAATGTCAGT 1500

Qy   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACG 1560

Qy   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCT 1620

Qy   1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          |||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db   1621 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCTAAATATCTTGCTG 1680

Qy   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
          |||||||||||| ||||||||||||||||||||  ||||||||||||||||||||||||
Db   1681 CAGAGAGAACGAGCCTTGGCATCTTTGGGAAAGCCCTTTCACCCTTTGCTGGTGGTTGCC 1740

Qy   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCCAACCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1800

Qy   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCA 1860

Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1920

Qy   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAA 1980

Qy   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGGATGGGGAAAGATGCC 2040
```

```
Qy    2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    2041 ACCCTCCTGATACATGAAGCCACCCTGGAAGACGGTTTGGAAGAGGAAGCAGTGGAAAAG 2100

Qy    2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATT 2160

Qy    2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    2161 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAAC 2220

Qy    2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    2221 GAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGAGACTTTGCAACAATG 2280

Qy    2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 CCCAAGCTGATTCCCCCACTGAAAGCCCTGTTTGCTGGCGACATCGAGGAGATGGAGGAG 2340

Qy    2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtccagggagctggca 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CGCAGGGAGAAGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCCAGGGAGCTGGCA 2400

Qy    2401 ggcggcctggaggatggggagcctcagcagaa 2432
           |||||||||||||||||||||||||||||||
Db    2401 GGCGGCCTGGAGGATGGGGAGCCTCAGCAGAA 2432


RESULT   6
AF308694
LOCUS       AF308694     2893 bp    mRNA            PRI       27-FEB-2001
DEFINITION  Gorilla gorilla ELAC2 (ELAC2) mRNA, complete cds.
ACCESSION   AF308694
VERSION     AF308694.1  GI:10946488
KEYWORDS    .
SOURCE      gorilla.
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
   PUBMED   11175785
REFERENCE   2  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Gorilla gorilla ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2893)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
FEATURES             Location/Qualifiers
     source          1. .2893
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
     gene            1. .2893
                     /gene="ELAC2"
     CDS             1. .2481
                     /gene="ELAC2"
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24916.1"
                     /db_xref="GI:10946489"
                     /translation="MWALCSLLRSAAGRTMSQGRTISQAPARRERPRKDPLRHLRTRE
                     KRGPSGCSGGPNTVYLQVVAAGSRDSGAALYVFSEFNRYLFNCGEGVQRLMQEHKLKV
                     VRLDNIFLTRMHWSNVGGLSGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKG
                     IELAVRPHSAPEYEDETMTVYQIPIHSEQRRGRHQPWQSPERPLSRLSPERSSDSESN
                     ENEPHLPHGVSQRRGVRDSSLVVAFICKLHLKRGNFLVLKAKEMGLPVGTAAIAPIIA
                     AVKDGKSITHEGREILAEELCTPPDPGAAFVVVECPDESFIQPICENATFQRYQGKAD
                     APVALVVHMAPESVLVDSRYQQWMERFGPDTQHLVLNENCASVHNLRSHKIQTQLNLI
                     HPDIFPLLTSFPCKKEGPTLSVPMVQGECLLKYQLRPRREWQRDAIITCNPEEFIVEA
                     LQLPNFQQSVQEYRRSVQDVPAPAEKRSQYPEIIFLGTGSAIPMKIRNVSATLVNISP
                     DTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADHHTGLLNILLQREQAL
                     ASLGKPLHPLLVVAPSQLKAWLQQYHNQCQEVLHHISMIPAKCLQEGAEISSPAVERL
                     ISSLLRTCDLEEFQTCLVRHCKHAFGCALVHTSGWKVVYSGDTMPCEALVRMGKDATL
                     LIHEATLEDGLEEEAVEKTHSTTSQAISVGMRMNAEFIMLNHFSQRYAKVPLFSPNFN
                     EKVGVAFDHMKVCFGDFPTMPKLIPPLKALFAGDIEEMEERREKRELRQVRAALLSGE
                     LAGGLEDGEPQQKRAHTEEPQAKKVRAQ"
BASE COUNT      705 a    787 c    815 g    586 t
ORIGIN


  Query Match             51.0%;  Score 1266;  DB 9;  Length 2893;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 2366;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 60

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 120

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    121 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGGGGCCCAAACACCGTGTACCTG 180

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 240

Qy    241 aaccggtatctcttcaactgtggagaaggcgttcagagactcatgcaggagcacaagtta 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AACCGGTATCTCTTCAACTGTGGAGAAGGCGTTCAGAGACTCATGCAGGAGCACAAGTTA 300

Qy    301 aaggttgctcgcctggacaacatattcctgacacgaatgcactggtctaatgttgggggc 360
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGGTTGTTCGCCTGGACAACATATTCCTGACACGAATGCACTGGTCTAATGTTGGGGGC 360

Qy    361 ttaagtggaatgattcttactttaaaggaaaccgggcttccaaagtgtgtactttctgga 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TTAAGTGGAATGATTCTTACTTTAAAGGAAACCGGGCTTCCAAAGTGTGTACTTTCTGGA 420

Qy    421 cctccacaactggaaaaatacctcgaagcaatcaaaatattttctggtccattgaaagga 480
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTCCACAGCTGGAAAAATACCTCGAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGA 480
```

```
Qy    481 atagaactggctgtgcggccccactctgccccagaatacgaggatgaaaccatgacagtt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATAGAACTGGCTGTGCGGCCCCACTCTGCCCCAGAATACGAGGATGAAACCATGACAGTT 540

Qy    541 taccagatccccatacacagtgaacagaggaggggaaagcaccaaccatggcagagtcca 600
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    541 TACCAGATCCCCATACACAGTGAACAGAGGAGGGGAAGGCACCAACCATGGCAGAGTCCA 600

Qy    601 gaaaggcctctcagcaggctcagtccagagcgatcttcagactccgagtcgaatgaaaat 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAAAGGCCTCTCAGCAGGCTCAGTCCAGAGCGATCTTCAGACTCCGAGTCGAATGAAAAT 660

Qy    661 gagccacaccttccacatggtgttagccagagaagaggggtcagggactcttccctggtc 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAGCCACACCTTCCACATGGTGTTAGCCAGAGAAGAGGGGTCAGGGACTCTTCCCTGGTC 720

Qy    721 gtagctttcatctgtaagcttcacttaaagagaggaaacttcttggtgctcaaagcaaag 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GTAGCTTTCATCTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGCTCAAAGCAAAG 780

Qy    781 gagatgggcctcccagttgggacagctgccatcgctcccatcattgctgctgtcaaggac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGATGGGCCTCCCAGTTGGGACAGCTGCCATCGCTCCCATCATTGCTGCTGTCAAGGAC 840

Qy    841 gggaaaagcatcactcatgaaggaagagagattttggctgaagagctgtgtactcctcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GGGAAAAGCATCACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCCTCCA 900

Qy    901 gatcctggtgctgcttttgtggtggtagaatgtccagatgaaagcttcattcaacccatc 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GATCCTGGTGCTGCTTTTGTGGTGGTAGAATGTCCAGATGAAAGCTTCATTCAACCCATC 960

Qy    961 tgtgagaatgccacctttcagaggtaccaaggaaaggcagatgcccccgtggccttggtg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TGTGAGAATGCCACCTTTCAGAGGTACCAAGGAAAGGCAGATGCCCCCGTGGCCTTGGTG 1020

Qy   1021 gttcacatggccccagcatctgtgcttgtggacagcaggtaccagcagtggatggagagg 1080
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCACATGGCCCCAGAATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGG 1080

Qy   1081 tttgggcctgacacccagcacttggtcctgaatgagaactgtgcctcagttcacaacctt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTGGGCCTGACACCCAGCACTTGGTCCTGAATGAGAACTGTGCCTCAGTTCACAACCTT 1140

Qy   1141 cgcagccacaagattcaaacccagctcaacctcatccacccggacatcttcccctgctc 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGCAGCCACAAGATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCCCTGCTC 1200

Qy   1201 accagtttccgctgtaagaaggagggccccacccctcagtgtgcccatggttcagggtgaa 1260
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAGTTTCCCCTGTAAGAAGGAGGGCCCCACCCTCAGTGTGCCCATGGTTCAGGGTGAA 1260

Qy   1261 tgcctcctcaagtaccagctccgtcccaggagggagtggcagagggatgccattattact 1320
          |||||||||||||||||||||||||||||||| |||||||||||||||||||| |||||||
Db   1261 TGCCTCCTCAAGTACCAGCTCCGTCCCAGGAGGGAATGGCAGAGGGATGCCATTATCACT 1320

Qy   1321 tgcaatcctgaggaattcatagttgaggcgctgcagcttcccaacttccagcagagcgtg 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1321 TGCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGTGTG 1380

Qy   1381 caggagtacaggaggagtgcgcaggacggcccagccccagcagagaaaagaagtcagtac 1440
          |||||||||||||||||| |||||||| ||||||||||||||||||||||||||||||||
Db   1381 CAGGAGTACAGGAGGAGTGTGCAGGACGTCCCAGCCCCAGCAGAGAAAAGAAGTCAGTAC 1440

Qy   1441 ccagaaatcatcttccttggaacagggtctgccatcccgatgaagattcgaaatgtcagt 1500
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   1441 CCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCCCATGAAGATTCGAAATGTCAGT 1500

Qy   1501 gccacacttgtcaacataagccccgacacgtctctgctactggactgtggtgagggcaca 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GCCACACTTGTCAACATAAGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACG 1560

Qy   1561 tttgggcagctgtgccgtcattacggagaccaggtggacagggtcctgggcaccctggct 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTTGGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCT 1620

Qy   1621 gctgtgtttgtgtcccacctgcacgcagatcaccacacgggcttgccaagtatcttgctg 1680
          |||||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db   1621 GCTGTGTTTGTGTCCCACCTGCACGCAGATCACCACACGGGCTTGCTAAATATCTTGCTG 1680

Qy   1681 cagagagaacgcgccttggcatctttgggaaagccgcttcacccttgctggtggttgcc 1740
          |||||||||| ||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   1681 CAGAGAGAACAAGCCTTGGCATCTTTGGGAAAGCCCCTTCACCCTTTGCTGGTGGTTGCC 1740

Qy   1741 cccaaccagctcaaagcctggctccagcagtaccacaaccagtgccaggaggtcctgcac 1800
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CCCAGCCAGCTCAAAGCCTGGCTCCAGCAGTACCACAACCAGTGCCAGGAGGTCCTGCAC 1800

Qy   1801 cacatcagtatgattcctgccaaatgccttcaggaaggggctgagatctccagtcctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 CACATCAGTATGATTCCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCCTGCA 1860

Qy   1861 gtggaaagattgatcagttcgctgttgcgaacatgtgatttggaagagtttcagacctgt 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GTGGAAAGATTGATCAGTTCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1920

Qy   1921 ctggtgcggcactgcaagcatgcgtttggctgtgcgctggtgcacacctctggctggaaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CTGGTGCGGCACTGCAAGCATGCGTTTGGCTGTGCGCTGGTGCACACCTCTGGCTGGAAA 1980

Qy   1981 gtggtctattccggggacaccatgccctgcgaggctctggtccggatggggaaagatgcc 2040
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   1981 GTGGTCTATTCCGGGGACACCATGCCCTGCGAGGCTCTGGTCCGCATGGGGAAAGATGCC 2040

Qy   2041 accctcctgatacatgaagccaccctggaagatggtttggaagaggaagcagtggaaaag 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 ACCCTCCTGATACATGAAGCCACCCTGGAAGATGGTTTGGAAGAGGAAGCAGTGGAAAAG 2100

Qy   2101 acacacagcacaacgtcccaagccatcagcgtggggatgcggatgaacgcggagttcatt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACACACAGCACAACGTCCCAAGCCATCAGCGTGGGGATGCGGATGAACGCGGAGTTCATT 2160

Qy   2161 atgctgaaccacttcagccagcgctatgccaaggtccccctcttcagccccaacttcagc 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2161 ATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACTTCAAC 2220

Qy   2221 gagaaagtgggagttgcctttgaccacatgaaggtctgctttggagactttccaacaatg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAGAAAGTGGGAGTTGCCTTTGACCACATGAAGGTCTGCTTTGGAGACTTTCCAACAATG 2280

Qy   2281 cccaagctgattcccccactgaaagccctgtttgctggcgacatcgaggagatggaggag 2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CCCAAGCTGATTCCCCCACTGAAAGCCCTGTTTGCCGGCGACATCGAGGAGATGGAGGAG 2340

Qy   2341 cgcagggagaagcgggagctgcggcaggtgcgggcggccctcctgtcc 2388
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CGCAGGGAGAAGCGGGAGCTGCGGCAGGTGCGGGCGGCCCTCCTGTCC 2388
```

```
RESULT   7
AC005277
LOCUS       AC005277   118788 bp    DNA             PRI       23-JUL-1998
DEFINITION  Homo sapiens chromosome 17, clone hRPK.597_M_12, complete sequence.
ACCESSION   AC005277
VERSION     AC005277.1  GI:3337311
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPK.597_M_12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 118788)
```

```
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 118788)
  AUTHORS   Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatin,C.,
            Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
            Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
            FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
            Gensheimer,S., Geraigery,K., Gilmartin,T., Grant,G., Hagos,B.,
            Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,
            Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
            Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,
            Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
            Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1998) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 23, 1998 this sequence version replaced gi:3335015.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Only the first 118.8 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC005274 (WICGR project
            L350).
FEATURES             Location/Qualifiers
     source          1. .118788
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="hRPK.597_M_12"
                     /clone_lib="RPCI-11 human BAC library"
                     /map="17"
                     /chromosome="17"
     repeat_region   795. .884
                     /rpt_family="MIR"
     repeat_region   complement(932. .1000)
                     /rpt_family="MIR"
     repeat_region   2044. .2076
                     /rpt_family="AT_rich"
     repeat_region   2237. .2713
                     /rpt_family="MER34"
     repeat_region   2718. .3477
                     /rpt_family="L2"
     repeat_region   complement(3478. .3821)
                     /rpt_family="MLT1A1"
     repeat_region   3822. .4007
                     /rpt_family="L2"
     repeat_region   complement(4015. .4281)
                     /rpt_family="MER89"
     repeat_region   complement(4282. .4321)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(4328. .4608)
                     /rpt_family="AluSx"
     repeat_region   complement(4609. .4895)
                     /rpt_family="MER89"
     repeat_region   complement(4896. .5186)
                     /rpt_family="AluSp"
     repeat_region   5219. .5479
                     /rpt_family="L2"
     repeat_region   complement(5481. .5740)
                     /rpt_family="MER34"
     repeat_region   complement(5963. .6275)
                     /rpt_family="AluSp"
     repeat_region   complement(6411. .6494)
                     /rpt_family="MER5A"
     repeat_region   6830. .7013
                     /rpt_family="MLT1C"
     repeat_region   7014. .7318
                     /rpt_family="AluYb8"
     repeat_region   7319. .7581
                     /rpt_family="MLT1C"
     repeat_region   7721. .8032
                     /rpt_family="AluJb"
     repeat_region   8039. .8333
                     /rpt_family="AluSq"
     repeat_region   9431. .9460
                     /rpt_family="AT_rich"
     repeat_region   9662. .9829
                     /rpt_family="MIR"
     repeat_region   complement(11302. .11728)
                     /rpt_family="MER39"
     repeat_region   complement(11824. .11884)
                     /rpt_family="LTR16C"
     repeat_region   complement(11928. .12108)
                     /rpt_family="LTR16C"
     repeat_region   12453. .12725
                     /rpt_family="AluJb"
     repeat_region   13377. .13785
                     /rpt_family="MSTB"
     repeat_region   14118. .14153
                     /rpt_family="(CA)n"
     repeat_region   complement(15320. .15620)
                     /rpt_family="AluJo"
     repeat_region   15683. .15759
                     /rpt_family="(CA)n"
     repeat_region   complement(15838. .15893)
                     /rpt_family="MER91A"
     repeat_region   complement(15915. .16080)
                     /rpt_family="MER91A"
     repeat_region   complement(16474. .16544)
                     /rpt_family="L2"
     repeat_region   17907. .17920
                     /rpt_family="L1ME"
     repeat_region   18271. .18311
                     /rpt_family="L2"
     repeat_region   complement(18368. .18529)
                     /rpt_family="AluJb"
     repeat_region   complement(18530. .18560)
                     /rpt_family="(CA)n"
     repeat_region   complement(18561. .18681)
                     /rpt_family="AluJb"
     repeat_region   18760. .19065
                     /rpt_family="AluSg"
     repeat_region   complement(19638. .19896)
                     /rpt_family="MIR"
     repeat_region   complement(19971. .20260)
                     /rpt_family="L1ME3A"
     repeat_region   20263. .20552
                     /rpt_family="MER58B"
     repeat_region   complement(20553. .20816)
                     /rpt_family="L1ME3A"
     repeat_region   complement(20821. .21069)
                     /rpt_family="L1M4"
     repeat_region   complement(21071. .21100)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(21101. .21338)
```

```
                /rpt_family="AluSx"
     repeat_region   complement(21348. .21598)
                /rpt_family="L1M4"
     repeat_region   21799. .22104
                /rpt_family="AluJo"
     repeat_region   22151. .22562
                /rpt_family="MSTA"
     repeat_region   23978. .24287
                /rpt_family="AluSq"
     repeat_region   24852. .25199
                /rpt_family="L2"
     repeat_region   complement(25697. .25985)
                /rpt_family="MLT1D"
     unsure          25816. .25895
                /note="Single-stranded coverage."
     repeat_region   complement(25994. .26099)
                /rpt_family="purine-rich"
     repeat_region   complement(26122. .26291)
                /rpt_family="MLT1D"
     repeat_region   26312. .26337
                /rpt_family="(CAAAA)n"
     repeat_region   26456. .26480
                /rpt_family="(GGGAA)n"
     repeat_region   complement(26576. .26781)
                /rpt_family="MER6B"
     repeat_region   complement(26912. .26990)
                /rpt_family="MER21B"
     repeat_region   26991. .27290
                /rpt_family="AluSx"
     repeat_region   complement(27291. .27975)
                /rpt_family="MER21B"
     repeat_region   28034. .28197
                /rpt_family="LTR28"
     repeat_region   28266. .28795
                /rpt_family="LTR28"
     repeat_region   28843. .28916
                /rpt_family="LTR28"
     repeat_region   28920. .29005
                /rpt_family="MIR"
     repeat_region   29140. .29233

  Query Match             10.0%;  Score 247;  DB 9;  Length 118788;
  Best Local Similarity   100.0%;  Pred. No. 2e-131;
  Matches  247;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 atgtgggcgctttgctcgctgctgcggtccgcggccggacgcaccatgtcgcagggacgc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78562 ATGTGGGCGCTTTGCTCGCTGCTGCGGTCCGCGGCCGGACGCACCATGTCGCAGGGACGC 78621

Qy     61 accatatcgcaggcacccgcccgccgcgagcggccgcgcaaggacccgctgcggcacctg 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78622 ACCATATCGCAGGCACCCGCCCGCCGCGAGCGGCCGCGCAAGGACCCGCTGCGGCACCTG 78681

Qy    121 cgcacgcgagagaagcgcggaccgtcggggtgctccggcggcccaaacaccgtgtacctg 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78682 CGCACGCGAGAGAAGCGCGGACCGTCGGGGTGCTCCGGCGGCCCAAACACCGTGTACCTG 78741

Qy    181 caggtggtggcagcgggtagccgggactcgggcgccgcgctctacgtcttctccgagttc 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78742 CAGGTGGTGGCAGCGGGTAGCCGGGACTCGGGCGCCGCGCTCTACGTCTTCTCCGAGTTC 78801

Qy    241 aaccggt 247
          |||||||
Db  78802 AACCGGT 78808


RESULT   8
AF304371S2
LOCUS       AF304371S2    740 bp    DNA             PRI       19-OCT-2000
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 gene, exon 17 and partial cds.
ACCESSION   AF304369
VERSION     AF304369.1  GI:10880929
KEYWORDS    .
SEGMENT     2 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 740)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1. .740
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     exon            301. .439
                     /number=17
     mRNA            <301. .>439
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
     CDS             <301. .>439
                     /codon_start=2
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24440.1"
                     /db_xref="GI:10880931"
                     /translation="PDTSLLLDCGEGTFGQLCRHYGDQVDRVLGTLAAVFVSHLHADH
                     HT"
BASE COUNT      154 a    188 c    173 g    225 t
ORIGIN


  Query Match             5.7%;  Score 142;  DB 9;  Length 740;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-70;
  Matches  142;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1519 agccccgacacgtctctgctactggactgtggtgagggcacatttgggcagctgtgccgt 1578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 AGCCCCGACACGTCTCTGCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTGTGCCGT 358

Qy   1579 cattacggagaccaggtggacagggtcctgggcaccctggctgctgtgtttgtgtcccac 1638
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 CATTACGGAGACCAGGTGGACAGGGTCCTGGGCACCCTGGCTGCTGTGTTTGTGTCCCAC 418

Qy   1639 ctgcacgcagatcaccacacgg 1660
          |||||||||||||||||||||
Db    419 CTGCACGCAGATCACCACACGG 440


RESULT   9
```

```
AF304371S1
LOCUS       AF304371S1    721 bp    DNA             PRI       19-OCT-2000
DEFINITION  Homo sapiens putative prostate cancer susceptibility protein
            HPC2/ELAC2 gene, exon 7 and partial cds.
ACCESSION   AF304371
VERSION     AF304371.1  GI:10880928
KEYWORDS    .
SEGMENT     1 of 2
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     A strong candidate prostate cancer susceptibility gene at
            chromosome 17p
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 721)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1. .721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p"
     exon            301. .420
                     /number=7
     mRNA            <301. .>420
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
     CDS             <301. .>420
                     /codon_start=3
                     /product="putative prostate cancer susceptibility protein
                     HPC2/ELAC2"
                     /protein_id="AAG24439.1"
                     /db_xref="GI:10880930"
                     /translation="EQRRGKHQPWQSPERPLSRLSPERSSDSESNENEPHLPH"
BASE COUNT      201 a    143 c    184 g    193 t
ORIGIN


  Query Match             4.9%;  Score 122;  DB 9;  Length 721;
  Best Local Similarity   100.0%;  Pred. No. 9.5e-59;
  Matches  122;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     560 gtgaacagaggaggggaaagcaccaaccatggcagagtccagaaaggcctctcagcaggc 619
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GTGAACAGAGGAGGGGAAAGCACCAACCATGGCAGAGTCCAGAAAGGCCTCTCAGCAGGC 360

Qy     620 tcagtccagagcgatcttcagactccgagtcgaatgaaaatgagccacaccttccacatg 679
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCAGTCCAGAGCGATCTTCAGACTCCGAGTCGAATGAAAATGAGCCACACCTTCCACATG 420

Qy     680 gt 681
           ||
Db     421 GT 422


RESULT  10
AF308696
LOCUS       AF308696     2712 bp    mRNA            ROD       27-FEB-2001
DEFINITION  Mus musculus ELAC2 mRNA, complete cds.
ACCESSION   AF308696
VERSION     AF308696.2  GI:11992378
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Abtin,V., Baumgard,M.,
            Beck,A., Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P.,
            Desrochers,M., Dumont,M., Farnham,J.M., Frank,D., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swedlund,B., Swensen,J.,
            Thomas,A., Tranchant,M., Woodland,A.M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     A candidate prostate cancer susceptibility gene at chromosome 17p
  JOURNAL   Nat. Genet. 27 (2), 172-180 (2001)
  PUBMED    11175785
REFERENCE   2  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Mouse ortholog of human HPC2/ELAC2
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
REFERENCE   4  (bases 1 to 2712)
  AUTHORS   Tavtigian,S.V., Simard,J., Teng,D.H.F., Baumgard,M., Beck,A.,
            Camp,N.J., Carillo,A.R., Chen,Y., Dayananth,P., Desrochers,M.,
            Dumont,M., Farnham,J.M., Frank,D., Frye,C., Ghaffari,S.,
            Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N., Laity,K.E.,
            Leavitt,A., Leblanc,G., McArthur-Morrison,J., Pederson,A., Penn,B.,
            Peterson,K.T., Reid,J.E., Richards,S., Schroeder,M., Smith,R.,
            Snyder,S.C., Swedlund,B., Swensen,J., Thomas,A., Tranchant,M.,
            Woodland,A.-M., Labrie,F., Skolnick,M.H., Neuhausen,S., Rommens,J.
            and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-DEC-2000) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84103, USA
  REMARK    Sequence update by submitter
COMMENT     On Dec 26, 2000 this sequence version replaced gi:10946492.
FEATURES             Location/Qualifiers
     source          1. .2712
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /dev_stage="fetus"
```

```
     CDS             15. .2510
                     /note="similar to Homo sapiens HPC2/ELAC2"
                     /codon_start=1
                     /product="ELAC2"
                     /protein_id="AAG24918.2"
                     /db_xref="GI:11992379"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPSKDPLRHLRTREKRGPGP
                     GGPNTVYLQVVAAGGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERRCGKQQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     FPQLTSFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTPDCNTDEFIAEALELP
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     LETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAGFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPLKALFAGDIEEMVERREKRELRLVRAALLTQQADSP
                     EDREPQQKRAHTDEPHSPQSKKESVANTLGARV"
BASE COUNT      659 a    741 c    748 g    564 t
ORIGIN


  Query Match             1.9%;  Score 47;  DB 10;  Length 2712;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-15;
  Matches   47;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    445 gaagcaatcaaaatattttctggtccattgaaaggaatagaactggc 491
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 GAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGGC 481


RESULT  11
AF348157
LOCUS       AF348157    34593 bp    DNA             ROD       04-APR-2001
DEFINITION  Mus musculus putative prostate cancer susceptibility protein
            (Elac2) gene, complete cds, alternatively spliced.
ACCESSION   AF348157
VERSION     AF348157.1  GI:13540341
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Mouse Elac2-containing genomic DNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 34593)
  AUTHORS   Frank,D.C., Swedlund,B., Dumont,M., Tavtigian,S.V., Simard,J.,
            Teng,D.H.F., Baumgard,M., Beck,A., Camp,N.J., Carillo,A.R.,
            Chen,Y., Dayananth,P., Desrochers,M., Farnham,J.M., Frye,C.,
            Ghaffari,S., Gupte,J.S., Hu,R., Iliev,D., Janecki,T., Kort,E.N.,
            Laity,K.E., Leavitt,A., Leblanc,G., McArthur-Morrison,J.,
            Pederson,A., Penn,B., Peterson,K.T., Reid,J.E., Richards,S.,
            Schroeder,M., Smith,R., Snyder,S.C., Swensen,J., Thomas,A.,
            Tranchant,M., Woodland,A.-M., Labrie,F., Skolnick,M.H.,
            Neuhausen,S., Rommens,J. and Cannon-Albright,L.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2001) Myriad Genetics, Inc., 320 Wakara Way, Salt
            Lake City, UT 84108, USA
FEATURES             Location/Qualifiers
     source          1. .34593
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     mRNA            join(<8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .>30639)
                     /gene="Elac2"
                     /note="alternatively spliced 24 exon form"
                     /product="putative prostate cancer susceptibility protein"
     mRNA            join(<8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .30624,
                     31242. .>31277)
                     /gene="Elac2"
                     /note="alternatively spliced 25 exon form"
                     /product="putative prostate cancer susceptibility protein"
     gene            <8470. .>31277
                     /gene="Elac2"
     CDS             join(8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .30639)
                     /gene="Elac2"
                     /note="24 exon form"
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAK29421.1"
                     /db_xref="GI:13540343"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTREKRGPGP
                     GGPNTVYLQVVAAGGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERRCGKQQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     FPQLTSFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTLDCNTDEFIAEALELP
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
                     KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
                     LETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
                     ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
                     IAFDHMKVCFGDFPTVPKLIPPLKALFAGDIEEMVERREKRELRLVRAALLTQQADSP
                     EDREPQQKRAHTDEPHSPQSKKVRAQ"
     CDS             join(8470. .8690,9077. .9127,9266. .9336,9948. .10012,
                     10957. .11014,12096. .12164,12933. .13064,14902. .14960,
                     16609. .16667,16746. .16818,17370. .17482,18671. .18766,
                     21539. .21677,21757. .21842,23553. .23671,24619. .24715,
                     27291. .27429,27763. .27801,27930. .28039,28553. .28652,
                     28733. .28853,29101. .29179,29710. .29854,30406. .30624,
                     31242. .31277)
                     /gene="Elac2"
                     /note="25 exon form"
                     /codon_start=1
                     /product="putative prostate cancer susceptibility protein"
                     /protein_id="AAK29420.1"
                     /db_xref="GI:13540342"
                     /translation="MWALRSLLRPLGLRTMSQGSARRPRPPKDPLRHLRTREKRGPGP
                     GGPNTVYLQVVAAGGRDAGAALYVFSEYNRYLFNCGEGVQRLMQEHKLKVARLDNIFL
                     TRMHWSNVGGLCGMILTLKETGLPKCVLSGPPQLEKYLEAIKIFSGPLKGIELAVRPH
                     SAPEYKDETMTVYQVPIHSERRCGKQQPSQSPRTSPNRLSPKQSSDSGSAENGQCPPE
                     DSSAGANRKAWGRDPSLVVAFVCKLHLRKGNFLVLKAKELGLPVGTAAIAPIIAAVKD
                     GKSITYEGREIAAEELCTPPDPGLVFIVVECPDEGFILPICENDTFKRYQAEADAPVA
                     LVVHIAPESVLIDSRYQQWMERFGPDTQHLILNENCPSVHNLRSHKIQTQLSLIHPDI
                     FPQLTSFYSKEEGSTLSVPTVRGECLLKYQLRPKREWQRDTTLDCNTDEFIAEALELP
                     SFQESVEEYRKNVQENPAPAEKRSQYPEIVFLGTGSAIPMKIRNVSSTLVNLSPDKSV
                     LLDCGEGTFGQLCRHYGQQIDRVLCSLTAVFVSHLHADHHTGLLNILLQREHALASLG
```

KPFQPLLVVAPTQLRAWLQQYHNHCQEILHHVSMIPAKCLQKGAEVSNTTLERLISLL
LETCDLEEFQTCLVRHCKHAFGCALVHSSGWKVVYSGDTMPCEALVQMGKDATLLIHE
ATLEDGLEEEAVEKTHSTTSQAINVGMRMNAEFIMLNHFSQRYAKIPLFSPDFNEKVG
IAFDHMKVCFGDFPTVPKLIPPLKALFAGDIEEMVERREKRELRLVRAALLTQQADSP
EDREPQQKRAHTDEPHSPQSKKESVANTLGARV"
BASE COUNT 8641 a 8030 c 8142 g 9780 t
ORIGIN

Query Match 1.9%; Score 46; DB 10; Length 34593;
Best Local Similarity 100.0%; Pred. No. 1.9e-14;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 445 gaagcaatcaaaatattttctggtccattgaaaggaatagaactgg 490
||||||||||||||||||||||||||||||||||||||||||||||
Db 10969 GAAGCAATCAAAATATTTTCTGGTCCATTGAAAGGAATAGAACTGG 11014

RESULT 12
AC005545/c
LOCUS AC005545 43514 bp DNA PRI 03-SEP-1998
DEFINITION Homo sapiens chromosome 19, cosmid R26634, complete sequence.
ACCESSION AC005545
VERSION AC005545.1 GI:3478638
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 43514)
AUTHORS Lamerdin,J.E., McCready,P.M., Skowronski,E., Adamson,A.W.,
Burkhart-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stilwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J.,
Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
Liu,S., Attix,C., Andreise,T., Trankheim,M., Amico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Kronmiller,B., Arellano,A., Montgomery,M., Ow,D., Nolan,M.,
Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
TITLE Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a
serine protease gene cluster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 43514)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE 3 (bases 1 to 43514)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT Map and sequence oriented from p telomere to centromere. Cosmid
R26634 overlaps cosmid R26660 (AC005328) to the left from bases 1
to 4,683 of this accession, and overlaps cosmid F8682 (AC005257) to
the right from bases 41,662 to 43,514. Additional chromosome 19 map
and sequence information may be obtained at:
http://www-bio.llnl.gov/bbrp/genome/genome.html.
FEATURES Location/Qualifiers
source 1. .43514
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R26634"
/chromosome="19"
/map="19p13.3 between CDC34 and D19S342"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
/note="Cosmid library constructed at LLNL from flow-sorted
chromosomes from hybrid 5HL2-B, which carries chromosome
19 as its only human chromosome."
misc_feature 46. .152
/note="DDS similarity to overlapping ESTs:
(46. .152) AA431731 zw77g03.s1 Soares testis NHT Homo
sapiens cDNA clone 782260 3'; (391. .340); 100%
identity.~(111. .152) AA609572 af15c02.s1 Soares testis NHT
Homo sapiens cDNA clone 1031714 3'; (372. .331); 100%
identity."
misc_feature complement(273. .608)
/note="DDS similarity to overlapping ESTs:
(608. .273) T05239 EST03128 Homo sapiens cDNA clone
HFBCT75. Score: 621 Identity: 325/337 (96%).~(718. .308)
AA496009 zv72a03.s1 Soares total fetus Nb2HF8 9w Homo
sapiens cDNA clone 759148 3'; Score: 755 Identity:
398/408 (97%)."
misc_feature 378. .707
/note="DDS similarity to overlapping ESTs:
AA431731 zw77g03.s1 Soares testis NHT Homo sapiens cDNA
clone 782260 3'; (339. .11); 99% identity.~AA609572
af15c02.s1 Soares testis NHT Homo sapiens cDNA clone
1031714 3'; (330. .5); 99% identity.~AA448181 zw83c05.s1
Soares testis NHT Homo sapiens cDNA clone 782792 3';
(336. .4); 99% identity.
"
CDS <378. .468
/note="Hypothetical human protein with no significant
similarity to known proteins"
/codon_start=2
/evidence=not_experimental
/product="R26660_2, partial CDS"
/protein_id="AAC34213.1"
/db_xref="GI:3478640"
/translation="LVSPALRCLEPPHLANLTLEDAAECLKQH"
misc_feature 1771. .3383
/note="DDS similarity to overlapping ESTs:
(3391. .3154) AA630776 ac14c10.s1 Stratagene HeLa cell s3
937216 Homo sapiens cDNA clone 856434 3' similar to
TR:G163328 G163328 LEUKEMIA VIRUS CELL RECEPTOR. ;
(262. .1); 99% identity.~(1771. .2449) AA167736 zq40d04.s1
Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
632167 3'; Score: 1207 Identity: 655/674
(97%).~(2109. .2638) AA843415 ak07f04.s1 Soares parathyroid
tumor NbHPA Homo sapiens cDNA clone IMAGE:1405279 3';
Score: 969 Identity: 516/547 (94%).~(2814. .2510) AA364678
EST75336 Pineal gland II Homo sapiens cDNA 5' end; Score:
582 Identity: 301/305 (98%).~(3200. .2766) H73448
yu02f02.r1 Homo sapiens cDNA clone 232635 5'. Score: 717
Identity: 413/446 (92%).~(3358. .2857) H28923 ym33c10.r1
Homo sapiens cDNA clone 49874 5'. Score: 756 Identity:
465/518 (89%).~(3383. .2950) W69851 zd48c04.r1 Soares fetal
heart NbHH19W Homo sapiens cDNA clone 343878 5' similar to
PIR:A45716 A45716 leukemia virus cell receptor ;
Score: 744 Identity: 418/440 (95%). ~Additional EST
matches:
AA617872, AA716429, W63772, AA305631, AA321888, AA371945,
R18826, T33092, T30887, F01584 and many others"
mRNA complement(2119. .>39837)
/note="Human AP-3 complex delta subunit mRNA"
/product="Homo sapiens delta-adaptin mRNA"
CDS complement(join(3296. .3391,9810. .9889,10209. .10330,
10996. .11081,11259. .11347,11830. .12019,12408. .12455,
12802. .12951,13983. .14054,14459. .14536,15248. .15425,
15871. .15944,16342. .16541,16661. .16736,17330. .17401,
17728. .17869,18345. .18490,19724. .19955,21985. .22215,
22286. .22434,22857. .23002,24481. .24529,24953. .25002,
28275. .28324,30213. .30244,31625. .31660,33594. .33701,
38134. .38214,38850. .38930,39742. .>39837))
/note="vesicle coat component, similar to alpha and gamma
adaptins"
/codon_start=1
/product="AP-3 complex delta subunit, partial CDS"
/protein_id="AAC34214.1"
/db_xref="GI:3522925"
/translation="AKYISQCIDEIKQELKQDNIAVKANAVCKLTYLQMLGYDISWAA
FNIIEVMSASKFTFKRIGYLAASQSFHEGTDVIMLTTNQIRKDLSSPSQYDTGVALTG
LSCFVTPDLARDLANDIMTLMSHTKPYIRKKALIEPLTNLIHSTSAMSLLYECVNTVI
AVLISLSSGMPNHSASIQLCVQKLRILIEDSDQNLKYLGLLAMSKILKTHPKSVQSHK

```
                     DLILQCLDDKDESIRLRALDLLYGMVSKKNLMEIVKKLMTHVDKAEGTTYRDELLTKI
                     IDICSQSNYQYITNFEWYISILVELTRLEGTRHGHLIAAQMLDVAIRVKAIRKFAVSQ
                     MSALLDSAHLLASSTQRNGICEVLYAAAWICGEFSEHLQEPHHTLEAMLRPRVTTLPG
                     HIQAVYVQNVVKLYASILQQKEQAGEAEGAQAVTQLMVDRLPQFVQSADLEVQERASC
                     ILQLVKHIQKLQAKDVPVAEEVSALFAGELNPVAPKAQKKVPVPEGLDLDAWINEPLS
                     DSESEDERPRAVFHEEEQRRPKHRPSEADEEELARRREARKQEQANNPFYIKSSPSPQ
                     KRYQDTPGVEHIPVVQIDLSVPLKVPGLPMSDQYVKLEEERRHRQKLEKDKRRKKRKE
                     KEKKGKRRHSSLPTESDEDIAPAQQVDIVTEEMPENALPSDEDDKDPNDPYRALDIDL
                     DKPLADSEKLPIQKHRNTETSKSPEKDVPMVEKKSKKPKKKEKKHKEKERDKEKKKEK
                     EKKAEDLDFWLSTTPPPAPAPAPAPVPSTDECEDAKTEAQGEEDDAEGQDQDKKSPKP
                     KKKKHRKEKEERTKGKKKSKKQPPGSEEAAGEPVQNGAPEEEQLPPESSYSLLAENSY
                     VKMTCDIRGSLQEDSQVTVAIVLENRSSSILKGMELSVLDSLNARMARPQGSSVHDGV
                     PVPFQLPPGVSNEAQYVFTIQSIVMAQKLKGTLSFIAKNDEGATHEKLDFRLHFSCSS
                     YLITTPCYSDAFAKLLESGDLSMSSIKVDGIRMSFQNLLAKICFHHHFSVVERVDSCA
                     SMYSRSIQGHHVCLLVKKGENSVSVDGKCSDSTLLSNLLEEMKATLAKC"
CDS                  complement(join(3296. .3391,9810. .9889,10209. .10330,
                     10996. .11081,11259. .11347,11830. .12019,12408. .12455,
                     12802. .12951,15248. .15425,15871. .15944,16342. .16541,
                     16661. .16736,17330. .17401,17728. .17869,18345. .18490,
                     19724. .19955,21985. .22215,22286. .22434,22857. .23002,
                     24481. .24529,24953. .25002,28275. .28324,30213. .30286,
                     30441. .30580,31531. .31660,33594. .33701,38134. .38214,
                     38850. .38930,39742. .>39837))
                     /function="vesicle coat component"
                     /function="similar to alpha and gamma adaptins"
                     /note="AP-3 complex delta subunit"
                     /codon_start=1
                     /product="delta-adaptin, partial CDS"
                     /protein_id="AAC34212.1"
                     /db_xref="GI:3478639"
                     /translation="AKYISQCIDEIKQELKQDNIAVKANAVCKLTYLQMLGYDISWAA
                     FNIIEVMSASKFTFKRIGYLAASQSFHEGTDVIMLTTNQIRKDLSSPSQYDTGVALTG
                     LSCFVTPDLARDLANDIMTLMSHTKPYIRKKAVLIMYKVFLKYPESLRPAFPRLKEKL
                     EDPDPGVQSAAVNVICELARRNPKNYLSLAPLFFKLMTSSTNNWVLIKIIKLFGALTP
                     LEPRLGKKLIEPLTNLIHSTSAMSLLYECVNTVIAVLISLSSGMPNHSASIQLCVQKL
                     RILIEDSDQNLKYLGLLAMSKILKTHPKSVQSHKDLILQCLDDKDESIRLRALDLLYG
                     MVSKKNLMEIVKKLMTHVDKAEGTTYRDELLTKIIDICSQSNYQYITNFEWYISILVE
                     LTRLEGTRHGHLIAAQMLDVAIRVKAIRKFAVSQMSALLDSAHLLASSTQRNGICEVL
                     YAAAWICGEFSEHLQEPHHTLEAMLRPRVTTLPGHIQAVYVQNVVKLYASILQQKEQA
                     GEAEGAQAVTQLMVDRLPQFVQSADLEVQERASCILQLVKHIQKLQAKDVPVAEEVSA
                     LFAGELNPVAPKAQKKVPVPEGLDLDAWINEPLSDSESEDERPRAVFHEEEQRRPKHR
                     PSEADEEELARRREARKQEQANNPFYIKSSPSPQKRYQDTPGVEHIPVVQIDLSVPLK
                     VPGLPMSDQYVKLEEERRHRQKLEKDKRRKKRKEKEKKGKRRHSSLPTESDEDIAPAQ
                     QVDIVTEEMPENALPSDEDDKDPNDPYRALDIDLDKPLADSEKLPIQKHRNTETSKSP
                     EKDVPMVEKKSKKPKKKEKKHKEKERDKEKKKEKEKKKSPKPKKKKHRKEKEERTKGK
                     KKSKKQPPGSEEAAGEPVQNGAPEEEQLPPESSYSLLAENSYVKMTCDIRGSLQEDSQ
                     VTVAIVLENRSSSILKGMELSVLDSLNARMARPQGSSVHDGVPVPFQLPPGVSNEAQY
                     VFTIQSIVMAQKLKGTLSFIAKNDEGATHEKLDFRLHFSCSSYLITTPCYSDAFAKLL
                     ESGDLSMSSIKVDGIRMSFQNLLAKICFHHHFSVVERVDSCASMYSRSIQGHHVCLLV
                     KKGENSVSVDGKCSDSTLLSNLLEEMKATLAKC"
repeat_region        3435. .3567
                     /rpt_family="AluSq"
repeat_region        3570. .3866
                     /rpt_family="AluY"
repeat_region        3867. .3906
                     /rpt_family="(TAAA)n"
repeat_region        3908. .4090
                     /rpt_family="AluSq"
repeat_region        4101. .4266
                     /rpt_family="FRAM"
repeat_region        complement(5793. .6096)
                     /rpt_family="AluJo"
repeat_region        complement(6348. .6855)
                     /rpt_family="MER9"
repeat_region        6938. .7227
                     /rpt_family="AluSg"
repeat_region        7255. .7402
                     /rpt_family="L1MB5"
repeat_region        7403. .7689
                     /rpt_family="AluSx"
repeat_region        7693. .8049
                     /rpt_family="L1MB5"
repeat_region        8105. .8399
                     /rpt_family="AluY"

  Query Match              0.9%;  Score 23;  DB 9;  Length 43514;
  Best Local Similarity   100.0%;  Pred. No. 0.46;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    653 atgaaaatgagccacaccttcca 675
          |||||||||||||||||||||||
Db  10491 ATGAAAATGAGCCACACCTTCCA 10469


RESULT  13
AF165142
LOCUS       AF165142   105490 bp    DNA             HTG       13-JUN-2001
DEFINITION  Homo sapiens chromosome 8 clone CTA-392C11 map 8p11-p12, WORKING
            DRAFT SEQUENCE, 1 ordered pieces.
ACCESSION   AF165142
VERSION     AF165142.1  GI:5499748
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105490)
  AUTHORS   Schilhabel,M.B., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
            Lehmann,R., Menzel,U., Polley,A., Reichwald,K., Schudy,A.,
            Siddiqui,R., Taudien,S., Wen,G., Schlegelberger,B., Siebert,R.,
            Rosenthal,A. and Platzer,M.
  TITLE     Chromosome 8 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105490)
  AUTHORS   Schudy,A., Blechschmidt,K., Schillhabel,M., Baumgart,C., Menzel,U.,
            Weber,J., Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-1999) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   105490: contig of 105490 bp in length.
FEATURES             Location/Qualifiers
     source          1. .105490
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8p11-p12"
                     /clone="CTA-392C11"
BASE COUNT    30720 a  21646 c  21259 g  31865 t
ORIGIN


  Query Match              0.9%;  Score 23;  DB 2;  Length 105490;
  Best Local Similarity   100.0%;  Pred. No. 0.47;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1428 aagaagtcagtacccagaaatca 1450
          |||||||||||||||||||||||
Db  39697 AAGAAGTCAGTACCCAGAAATCA 39719


RESULT  14
AC091560
LOCUS       AC091560   190195 bp    DNA             HTG       07-JUL-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-104D16 map 8, WORKING DRAFT
            SEQUENCE, 2 ordered pieces.
ACCESSION   AC091560
VERSION     AC091560.2  GI:14626327
```

```
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190195)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-104D16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190195)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 7, 2001 this sequence version replaced gi:13959165.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11660
                Center clone name: 104_D_16
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 189800 bases at least Q40
                Consensus quality: 189993 bases at least Q30
                Consensus quality: 190051 bases at least Q20
                Insert size: 192000; agarose-fp
                Insert size: 190095; sum-of-contigs
                Quality coverage: 11.0 in Q20 bases; agarose-fp
                Quality coverage: 11.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    66263: contig of 66263 bp in length
            *   66264 66363: gap of      100 bp
            *   66364   190195: contig of 123832 bp in length.
FEATURES            Location/Qualifiers
     source         1. .190195
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="8"
                    /map="8"
                    /clone="RP11-104D16"
                    /clone_lib="RPCI-11 Human Male BAC"
     misc_feature   1. .66263
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   66364. .190195
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
BASE COUNT    57071 a  36687 c  36123 g  60205 t    109 others
ORIGIN


  Query Match             0.9%;  Score 23;  DB 2;  Length 190195;
  Best Local Similarity   100.0%;  Pred. No. 0.48;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1428 aagaagtcagtacccagaaatca 1450
          |||||||||||||||||||||||
Db 161058 AAGAAGTCAGTACCCAGAAATCA 161080


RESULT  15
AC019331
LOCUS       AC019331   190705 bp    DNA             HTG       07-JUL-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-104D16, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC019331
VERSION     AC019331.4  GI:8569783
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190705)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190705)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7717162.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0104D16
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 49%
            Sequencing vector: plasmid; 51%
            Chemistry: Dye-primer ET; 49% of reads
            Chemistry: Dye-terminator Big Dye; 51% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 182645 bases at least Q40
            Consensus quality: 185230 bases at least Q30
            Consensus quality: 186612 bases at least Q20
            Insert size: 199000; agarose-fp
            Insert size: 189205; sum-of-contigs
            Quality coverage: 4.52 in Q20 bases; agarose-fp
            Quality coverage: 4.79 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2064: contig of 2064 bp in length
            *     2065     2164: gap of unknown length
            *     2165     3805: contig of 1641 bp in length
            *     3806     3905: gap of unknown length
            *     3906     8649: contig of 4744 bp in length
            *     8650     8749: gap of unknown length
            *     8750    13321: contig of 4572 bp in length
            *    13322    13421: gap of unknown length
            *    13422    19509: contig of 6088 bp in length
            *    19510    19609: gap of unknown length
            *    19610    24715: contig of 5106 bp in length
            *    24716    24815: gap of unknown length
            *    24816    30643: contig of 5828 bp in length
            *    30644    30743: gap of unknown length
            *    30744    40713: contig of 9970 bp in length
            *    40714    40813: gap of unknown length
            *    40814    48495: contig of 7682 bp in length
            *    48496    48595: gap of unknown length
            *    48596    57063: contig of 8468 bp in length
            *    57064    57163: gap of unknown length
            *    57164    67250: contig of 10087 bp in length
            *    67251    67350: gap of unknown length
            *    67351    82559: contig of 15209 bp in length
            *    82560    82659: gap of unknown length
            *    82660   102558: contig of 19899 bp in length
            *   102559   102658: gap of unknown length
            *   102659   124552: contig of 21894 bp in length
            *   124553   124652: gap of unknown length
            *   124653   152228: contig of 27576 bp in length
            *   152229   152328: gap of unknown length
            *   152329   190705: contig of 38377 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190705
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-104D16"
     misc_feature    1. .2064
                     /note="assembly_name:Contig7"
     misc_feature    2165. .3805
                     /note="assembly_name:Contig8
                     clone_end:SP6
                     vector_side:left"
     misc_feature    3906. .8649
                     /note="assembly_name:Contig9"
     misc_feature    8750. .13321
                     /note="assembly_name:Contig10"
     misc_feature    13422. .19509
                     /note="assembly_name:Contig11"
     misc_feature    19610. .24715
                     /note="assembly_name:Contig12"
     misc_feature    24816. .30643
                     /note="assembly_name:Contig13"
     misc_feature    30744. .40713
                     /note="assembly_name:Contig14"
     misc_feature    40814. .48495
                     /note="assembly_name:Contig15"
     misc_feature    48596. .57063
                     /note="assembly_name:Contig16"
     misc_feature    57164. .67250
                     /note="assembly_name:Contig17"
     misc_feature    67351. .82559
                     /note="assembly_name:Contig18"
     misc_feature    82660. .102558
                     /note="assembly_name:Contig19"
     misc_feature    102659. .124552
                     /note="assembly_name:Contig20"
     misc_feature    124653. .152228
                     /note="assembly_name:Contig21"
     misc_feature    152329. .190705
                     /note="assembly_name:Contig22"
BASE COUNT    55947 a  36657 c  35874 g  60716 t   1511 others
ORIGIN


  Query Match              0.9%;  Score 23;  DB 2;  Length 190705;
  Best Local Similarity   100.0%;  Pred. No. 0.48;
  Matches   23;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1428 aagaagtcagtacccagaaatca 1450
          |||||||||||||||||||||||
Db 161565 AAGAAGTCAGTACCCAGAAATCA 161587


Search completed: April  1, 2002, 11:18:05
Job time: 6329 sec
```